



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|   |  |  |  |
|---|--|--|--|
| (51) International Patent Classification <sup>6</sup> :<br><b>C12N 15/12, 15/10, 15/85, 5/10, C07K 14/51, C12Q 1/68</b>                   |  | A1   | (11) International Publication Number: <b>WO 95/33831</b>        |
|   |  |  | (43) International Publication Date: 14 December 1995 (14.12.95) |
| (21) International Application Number: PCT/US95/07349   |  | (81) Designated States: AM, AU, BB, BG, BR, BY, CA, CN, CZ, EE, FI, GE, HU, IS, JP, KG, KP, KR, KZ, LK, LR, LT, LV, MD, MG, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TT, UA, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, MW, SD, SZ, UG). |  |
| (22) International Filing Date: 7 June 1995 (07.06.95)  |  |  |  |
| (30) Priority Data:<br>255,250 7 June 1994 (07.06.94) US  |  |  |  |
| (71) Applicant: CREATIVE BIOMOLECULES, INC. [US/US]; 45 South Street, Hopkinton, MA 01748 (US).   |  | Published<br><i>With international search report.<br/>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>   |  |
| (72) Inventors: OZKAYNAK, Engin; 44 Purdue Drive, Milford, MA 01757 (US). OPPERMANN, Hermann; 25 Summer Hill Road, Medway, MA 02053 (US). |  |  |  |
| (74) Agent: KELLEY, Robin, D.; Testa, Hurwitz & Thibeault, High Street Tower, 125 High Street, Boston, MA 02110-2711 (US).                |  |  |  |

(54) Title: METHODS AND COMPOSITIONS FOR MODULATING MORPHOGENIC PROTEIN EXPRESSION

(57) Abstract

Disclosed are methods and compositions for screening compounds for their ability to modulate expression of a tissue morphogenetic protein, particularly OP-1, OP-1 homologs and closely related proteins, using one or more OP-1-specific, noncoding nucleotide sequences and a suitable reporter gene.

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |  |    |                          |
|----|--------------------------|----|--|----|--------------------------|
| AT | Austria                  | GB | United Kingdom                           | MR | Mauritania               |
| AU | Australia                | GE | Georgia                                  | MW | Malawi                   |
| BB | Barbados                 | GN | Guinea                                   | NE | Niger                    |
| BE | Belgium                  | GR | Greece                                   | NL | Netherlands              |
| BF | Burkina Faso             | HU | Hungary                                  | NO | Norway                   |
| BG | Bulgaria                 | IE | Ireland                                  | NZ | New Zealand              |
| BJ | Benin                    | IT | Italy                                    | PL | Poland                   |
| BR | Brazil                   | JP | Japan                                    | PT | Portugal                 |
| BY | Belarus                  | KE | Kenya                                    | RO | Romania                  |
| CA | Canada                   | KG | Kyrgyzstan                               | RU | Russian Federation       |
| CF | Central African Republic | KP | Democratic People's Republic<br>of Korea | SD | Sudan                    |
| CG | Congo                    | KR | Republic of Korea                        | SE | Sweden                   |
| CH | Switzerland              | KZ | Kazakhstan                               | SI | Slovenia                 |
| CI | Côte d'Ivoire            | LI | Liechtenstein                            | SK | Slovakia                 |
| CM | Cameroon                 | LK | Sri Lanka                                | SN | Senegal                  |
| CN | China                    | LU | Luxembourg                               | TD | Chad                     |
| CS | Czechoslovakia           | LV | Latvia                                   | TG | Togo                     |
| CZ | Czech Republic           | MC | Monaco                                   | TJ | Tajikistan               |
| DE | Germany                  | MD | Republic of Moldova                      | TT | Trinidad and Tobago      |
| DK | Denmark                  | MG | Madagascar                               | UA | Ukraine                  |
| ES | Spain                    | ML | Mali                                     | US | United States of America |
| FI | Finland                  | MN | Mongolia                                 | UZ | Uzbekistan               |
| FR | France                   |    |  | VN | Viet Nam                 |
| GA | Gabon                    |    |  |    |                          |

Methods and Compositions for Modulating  
Morphogenic Protein Expression

Reference to Related Applications

This application is a continuation-in-part of USSN 08/255,250, filed June 7, 1994 which is a continuation-in-part of USSN 07/938,021, filed August 28, 1992 which is a continuation-in-part of USSN 07/752,861, filed August 30, 1991 which is a continuation-in-part of USSN 07/667,274, filed March 11, 1991, the disclosures of which are incorporated herein by reference.

Field of the Invention

10 The invention relates generally to the field of drug screening assays. More particularly, the invention relates to methods and compositions for identifying molecules that modulate production of true tissue morphogenic proteins.

15 Background of the Invention

A class of proteins recently has been identified, the members of which are true tissue morphogenic proteins. The members of this class of proteins are characterized as competent for inducing the developmental cascade of cellular and molecular events that culminate in the formation of new organ-specific tissue, including any vascular and connective tissue formation as required by the naturally occurring tissue. Specifically, the morphogens are competent for inducing all of the following biological functions in a morphogenically permissive environment: (1) stimulating proliferation of progenitor cells; (2) stimulating differentiation of progenitor cells; (3) stimulating the proliferation of differentiated cells and (4) supporting the growth and maintenance of differentiated cells. For example, the morphogenic proteins can induce the full developmental cascade of bone tissue morphogenesis, including the migration and proliferation of mesenchymal cells, proliferation and differentiation of chondrocytes, cartilage matrix formation and calcification, vascular invasion, osteoblast proliferation, bone formation, bone remodeling, and hematopoietic bone marrow differentiation. These

proteins also have been shown to induce true tissue morphogenesis of non-chondrogenic tissue, including dentin, liver, and nerve tissue.

A particularly useful tissue morphogenic protein is human OP-1 (Osteogenic Protein-1), described in U.S. 5,011,691; US Pat. No. 5,266,683 and Ozkaynak et al. (1990) EMBO J. 9: 2085-2093. Species homologues identified to date include mouse OP-1 (see US Pat. 5,266,683) and the *Drosophila* homologue 60A, described in Wharton et al. (1991) PNAS 88:9214-9218). Other closely related 10 proteins include OP-2 (Ozkaynak (1992) J. Biol. Chem. 267:25220-25227 and US Pat. No. 5,266,683); BMP5, BMP6 (Celeste et al. (1991) PNAS 87:9843-9847) and Vgr-1 (Lyons et al. (1989)). These disclosures are incorporated herein by reference.

It previously has been contemplated that these tissue 15 morphogens can be administered to an animal to regenerate lost or damaged tissue. Alternatively, one can envision administering a molecule capable of modulating expression of the endogenous tissue morphogen as a means for providing morphogen to a site *in vivo*.

It is an object of this invention to provide compositions and 20 methods of screening compounds which can modulate expression of an endogenous tissue morphogen, particularly OP-1 and closely related genes. The compounds thus identified have utility both *in vitro* and *in vivo*. Useful compounds contemplated include those capable of stimulating transcription and/or translation of the OP-1 gene, 25 as well as compounds capable of inhibiting transcription and/or translation of the OP-1 gene.

These and other objects and features of the invention will be apparent from the description, drawings and claims which follow.

30 Summary of the Invention

The invention features compositions and methods for screening candidate compounds for the ability to modulate the effective local or systemic quantity of endogenous OP-1 in an organism, and methods for producing the compounds identified. In one aspect, 35 the method is practiced by: (1) incubating one or more candidate compounds with cells transfected with a DNA sequence encoding, in operative association with reporter gene, a portion of an OP-1

non-coding DNA sequence that is competent to act on and affect expression of the associated receptor gene; (2) measuring the level of reporter gene expression in the transfected cell, and (3) comparing the level of reporter gene expressed in the presence of 5 the candidate compound with the level of reporter gene expressed in the absence of the candidate compound. In a related aspect, the invention features the compound that is identified by use of the method of the invention.

The screening method of the invention provides a simple method 10 of determining a change in the level of a reporter gene product expressed by a cell following exposure to one or more compound(s). The level of an expressed reporter gene product in a given cell culture, or a change in that level resulting from exposure to one or more compound(s) indicates that application of the compound can 15 modulate the level of the morphogen expressed and normally associated with the non-coding sequence. Specifically, an increase in the level of reporter gene expression is indicative of a candidate compound's ability to increase OP-1 expression in vivo. Similarly, a decrease in the level of reporter gene 20 expression is indicative of a candidate compound's ability to decrease or otherwise interfere with OP-1 expression in vivo.

The methods and compositions of the invention can be used to identify compounds showing promise as therapeutics for various *in vivo* and *ex vivo* mammalian applications, as well as to identify 25 compounds having numerous utilities. For example, morphogen expression inducing compounds can be used *in vivo* to correct or alleviate a diseased condition, to regenerate lost or damaged tissue, to induce cell proliferation and differentiation, and/or to maintain cell and tissue viability and/or a differentiated 30 phenotype *in vivo* or *ex vivo*. The compounds also can be used to maintain the viability of, and the differentiated phenotype of, cells in culture. The various *in vivo*, *ex vivo*, and *in vitro* utilities and applications of the morphogenic proteins described herein are well documented in the art. See, for example, US 35 92/01968 (WO 94/03200), filed March 11, 1992; US 92/07358 (WO 93/04692), filed August 28; PCT US 92/0743 (WO 93/05751), filed August 28, 1992; US 93/07321 (WO 94/03200), filed July 29, 1993;

US 93/08808 (WO 94/06449), filed September 16, 1993; US93/08885 (WO94/06420), filed September 15, 1993, and US Pat. No. 5,266,683.

5 Morphogen expression inhibiting compounds identified by the methods, kits and compositions described herein can be used to modulate the degree and/or timing of morphogen expression in a cell. Such compounds can be used both *in vitro* and *in vivo* to more closely regulate the production and/or available concentration of morphogen.

10

List of useful terms and Definitions

As used herein, "gene expression" is understood to refer to the production of the protein product encoded by a DNA sequence of interest, including the transcription of the DNA sequence and 15 translation of the mRNA transcript.

As used herein, "operative association" is a fusion of the described DNA sequences with a reporter gene in such a reading frame as to be co-transcribed, or at such a relative positioning as to be competent to modulate expression of the reporter gene.

20 As used herein, "vector" is understood to mean any nucleic acid comprising a nucleotide sequence of interest and competent to be incorporated into a host cell and recombining with and integrating into the host cell genome. Such vectors include linear nucleic acids, plasmids, phagemids, cosmids, YAC'S (yeast 25 artificial chromosomes) and the like.

As used herein, "non-coding sequence" or "non-coding DNA" includes DNA sequences that are not transcribed into RNA sequence, and/or RNA sequences that are not translated into protein. This category of "non-coding sequence" has been defined for ease of 30 reference in the application, and includes sequences occurring 5' to the ATG site which indicates the start codon and sequences 3' to the stop codon, as well as intervening intron sequences that occur within the coding region of the gene. As used herein, an "OP1-specific" non-coding sequence is understood to define a non-coding sequence that lies contiguous to OP1 specific coding 35 sequence at an OP-1 gene locus under naturally-occurring

conditions. The sequences may include 5', 3' and intron sequences.

As used herein, "allelic, species and other sequence variants thereof" includes point mutations, insertions and deletions such 5 as would be naturally occurring or which can genetically engineered into an OP-1 non-coding DNA sequence and which do not affect substantially the regulation of a reporter gene by the OP-1 non-coding sequence. For example, one of ordinary skill in the art can use site directed mutagenesis to modify, as by deletion, 10 for example, one or more of the OP-1 non-coding sequences described herein without substantially affecting the regulation of OP-1 or a reporter gene by the modification. Such modifications are considered to be within the scope of the disclosure provided herein.

15 As used herein, a "Wt-1/Egr-1 consensus binding sequence" or Wt-1/Egr-1 consensus binding element" is a nine base sequence which has been shown to be bound by the DNA binding proteins Wt-1 and Egr-1. The consensus sequence of the Wt-1/Egr-1 binding site has been determined by homology to be GNGNGGGNG, Seq. ID No. 4 20 (Rauscher *et al.*, Science 250:1259-1262 (1990), incorporated herein by reference).

25 As used herein, a "TCC binding sequence" or "TCC binding element" is an approximately 15 to 20 base sequence of DNA which contains at least three contiguous or non-contiguous repeats of the DNA sequence TCC. The TCC binding sequence identified in human OP-1 genomic DNA is shown in Seq. ID No. 5, and the TCC binding sequence identified in murine OP-1 genomic DNA is shown in Seq. ID No. 6. The TCC binding sequence has also been shown to be bound by the DNA binding proteins Wt-1 and Egr-1 (Wang *et al.*, 30 Proc. Natl. Acad. Sci. 90:8896-8900 (1993); Wang *et al.*, Biochem. Biophys. Res. Comm., 188:433-439 (1992)).

35 As used herein, a "FTZ binding sequence" or "FTZ binding element" is a Fushi-tarazu DNA sequence (FTZ) that has been shown to be bound by the DNA binding protein Fushi-tarazu (FTZ-F1). The FTZ binding sequence identified in human OP-1 genomic DNA is shown in Seq. ID No. 7. The FTZ consensus sequence, a consensus sequence for the nuclear hormone receptor superfamily, is YCAAGGYCR.

As used herein, a "steroid binding sequence" or "steroid binding element" is a DNA sequence that has been shown to be bound by one or more elements, in response to activating signal molecules. Examples of such "activating signal molecules" include 5 retinoids, Vitamin D, and also include steroids such as estrogen and progesterone. Useful elements are anticipated to include the FTZ-F1 protein, WT-1 and Egr-1. Activating signal molecules of the nuclear receptor family have recently been shown to bind to DNA as homodimers, heterodimers or as monomers (Parker, M.G., 10 Curr. Op. Cell Biol., 1993, 5:499-504). The formation of heterodimers among the nuclear receptor family molecules may significantly increase the diversity of binding elements which are recognized by these nuclear receptors, and provide for differential regulation of genes containing the specific binding 15 sites. In addition, the nuclear receptors have been shown to interact with other accessory factors, such as transcription factors, to stimulate or repress transcription. These interactions, between the nuclear receptors and the nuclear receptors and accessory factors, indicate that there could be 20 significant number of nuclear receptor/accessory factor interactions which have widely different transcriptional activities.

While the method of the invention is described with reference to a single cell, as will be appreciated by those having ordinary 25 skill in the art, this is only for ease of description, and the method is most efficiently carried out using a plurality of cells.

With respect to transfection of DNA sequences in the cell and the method of the invention, all means for introducing nucleic acids into a cell are contemplated including, without limitation, 30 CaPO<sub>4</sub> co-precipitation, electroporation, DEAE-dextran mediated uptake, protoplast fusion, microinjection and lipofusion. A key to the invention is the DNA sequences with which the cell is transfected, rather than the mechanical or chemical process by which the DNA incorporation is accomplished.

35 Useful reporter genes are characterized as being easy to transfect into a suitable host cell, easy to detect using an established assay protocol, and genes whose expression can be tightly regulated. Other reporter genes contemplated to have

utility include, without limitation, the luciferase gene, the Green Fluorescent Protein (GFP) gene, the chloramphenicol Acetyl Transferase gene (CAT), human growth hormone, and beta-galactosidase. Additional useful reporter genes are any well characterized genes the expression of which is readily assayed, and examples of such reporter genes can be found in, for example, F.A. Ausubel et al., Eds., Current Protocols in Molecular Biology, John Wiley & Sons, New York, (1989). As will be appreciated by those having ordinary skill in the art, the listed reporter genes are only a few of the possible reporter genes, and it is only for ease of description that all available reporter genes are not listed.

While the method, vectors, and cells described recite the use of a reporter gene in operative association with an OP-1 non-coding DNA sequence, it will be apparent to those of ordinary skill in the art that the DNA sequence OP-1, including human OP1, shown in Seq. ID No. 1 or murine OP-1, disclosed in U.S. Patent No. 5,266,683, is also within the scope of a suitable reporter gene. Other suitable reporter genes can be used for ease in assaying for the presence of the reporter mRNA or reporter gene product.

Where a cell line is to be established, particularly where the transfected DNA is to be incorporated into the cell's genome, lines that can be immortalized are especially desirable. As used herein, "immortalized" cell lines are viable for multiple passages (e.g., greater than 50 generations) without significant reduction in growth rate or protein production.

While the selected non-coding DNA sequences disclosed herein are described using defined bases, as will be appreciated by those having ordinary skill in the art, to some degree the lengths of the selected DNA sequences recited are arbitrary and are defined for convenience. As will be understood by those of ordinary skill in the art, shorter sequences of OP-1 non-coding DNA sequence and other fusion DNA's can be used in a vector according to the invention, and can be transfected into a cell, or used in the method of the invention for screening a candidate compound for its ability to modulate OP-1 expression. Specifically, it is standard procedure for molecular biologists to first identify useful

regulatory sequences, and then to determine the minimum sequence required, by systematic digestion and mutagenesis e.g., by exonuclease or endonuclease digestion, site directed mutagenesis and the like. Accordingly, subsequent, standard routine 5 experimentation is anticipated to identify minimum sequences and these, shorter sequences are contemplated by the invention disclosed herein.

Useful cell types for the method and compositions according to the invention include any eukaryotic cell. Currently preferred 10 are cell types known to express OP-1. Such cells include epithelial cells and cells of uro-genital cell origin, including renal (kidney or bladder) cells, as well as liver, bone, nerve, ovary, cardiac muscle and the like. The cells may be derived from tissue or cultured from an established cell line. See, for example 15 Ozkaynak et al. (1991) Biochem. BioPhys. Res. Comm. 179:116-123 for a detailed description of tissues known to express OP-1. Other useful cells include those known to exhibit a steroid receptor, including cells having an estrogen receptor and cells responsive to the FTZ-F1 protein. Currently preferred cells also 20 have simple media component requirements. Other useful representative cells include, but are not limited to, Chinese hamster ovary (CHO); canine kidney (MDCK); or rat bladder (NBT-2), and the like. Useful cell types can be obtained from the American Type Culture Collection (ATCC), Rockville, MD or from the European 25 Collection of Animal Cell Cultures, Porton Down, Salisbury SP40JG, U.K. As used herein, "derived" means the cells are from the cultured tissue itself, or are a cell line whose parent cells are of the tissue itself.

30 Aspects and Embodiments of the Invention

In one aspect, the invention features a vector having a reporter gene operatively associated with a portion of one or more OP-1 non-coding sequences. The OP-1 non-coding sequence chosen is independently selected from the 5' (or "upstream") non-coding 35 human or murine OP-1 sequence shown in Seq. ID Nos. 1 and 2, respectively, the 3' (or "downstream") non-coding human or murine OP-1 sequence shown in Seq. ID Nos. 1 or 3, and the human intron non-coding OP-1 sequences shown in Seq. ID No. 1. Also

anticipated to be useful are the non-coding sequences (e.g., 5', 3' and intron) of other species homologs of OP-1 and proteins closely related to OP-1. In addition, the portion of OP-1 sequence included in the vector can be a combination of two or 5 more 5' non-coding, 3' non-coding and/or intron OP-1 sequences.

In one embodiment, the vector can include a non-coding OP1-specific sequence selected from at least one of the following sequence segments of Seq. ID No. 1 presented below, and which define human genomic OP-1 sequence comprising approximately 3.3 Kb 10 of 5' non-coding sequence. In Seq. ID No. 1, the start codon begins at position 3318, and the upstream sequence (bases 1 to 3317) is composed of untranscribed (1 to 2790) and untranslated (2791 to 3317) OP1-specific DNA; approximately 1 Kb of which is presented in Fig. 1 (bottom strand).

15 Useful sequence segments include bases 2548-3317, representing 750 bases sharing significant (greater than 70% identity) between the mouse and human OP-1 homologs (See Fig. 1), and bases 3170-3317; 3020-3317; 2790-3317; 2548-2790 of Seq. ID No. 1, all shorter fragments of this region of the DNA. As base 2790 is the 20 mRNA start site, other useful sequences include 2790-3317, representing transcribed but not translated 5' coding sequence and shorter fragments of this DNA region as noted above; upstream fragments of OP1-specific DNA, bases 2548-2790; 1549-2790; 1-2790 of Seq. ID No. 1. Also useful sequence segments include the 25 approximately 750 bases that have homology between the human and mouse OP-1 sequences with additional upstream sequences, 2300 to 3317; 1300 to 3317; 1-3317; all fragments of the disclosed upstream OP1-specific DNA sequences of Seq. ID No. 1.

In another embodiment, the sequences are defined by the non- 30 coding sequences of the mouse OP-1 homolog, including the following 5' non-coding sequences (Seq. ID No. 2): 2150-2296, 2000-2296, 1788-2296, and 1549-2296 all of which define the 750 bases sharing high sequence identity with the human homolog (See, Fig. 1); 800-2296; 1-2296; 1549-1788, 800-1788 and 1-1788.

35 Within this region also exist a number Egr/Wt-1 sites (8 in hOP-1; 7 in mOP-1), known in the art to bind the regulatory elements Egr and Wt-1. Accordingly, in another aspect, the invention contemplates a screening material for identifying

compounds which modulate OP-1 expression, the assay comprising the step of identifying compounds which bind Egr/Wt-1 site. At least one Wt/Egr-1 element, preferably between 1-6 elements, or at least 6 Wt/Egr-1 elements are included in a sequence. The relative 5 locations of these elements are indicated in Fig. 1 and at positions 3192-3200; 3143-3151; 3027-3035; 2956-2964; 2732-2740; 2697-2704 of Seq. ID No. 1, and positions 2003-2011; 1913-1922; 1818-1826; 1765-1776; 1757-1765; 1731-1739; 1699-1707; 1417-1425 of Seq. ID No. 2 of Seq. ID Nos. 1, 2 substantially the same Seq. 10 alignment. The lengths of bases within these 5' non-coding sequences is selected to include portions of the sequence of DNA which was determined to be homologous between murine and human genomic OP-1, separately and as a part of a larger sequence including non-homologous DNA. Additionally, the portion of OP-1 15 sequence selected can be a portion of the region of homology between murine and human OP-1 DNA sequences, bases 2548-2790 or 2548-3317 of Seq. ID No. 1, or bases 1549 to 1788 or 1549 to 2296 of Seq. ID No. 2, and/or at least one of an Wt-1/Egr-1 consensus binding sequence. In still another aspect the portion of OP-1 20 sequence selected can include a TCC binding sequence, a FTZ binding sequence, a steroid binding sequence, or part or all of an OP-1 intron sequence. The relative positions of the TCC and FTZ elements are indicated in Fig. 1 and at positions 2758-2778 (TCC); 2432-2441 (FTZ) of Seq. ID No. 1 and 1755-1769 (TCC) of Seq. ID 25 No. 2.

In another aspect, the invention features a cell that has been transfected with a reporter gene in operative association with a portion of OP-1 non-coding DNA sequence. The portion of OP-1 non-coding sequence is independently selected from the 5' (or 30 upstream) non-coding human or murine OP-1 sequence shown in Seq. ID Nos. 1 and 2, the 3' (or downstream) non-coding murine OP-1 sequence shown in Seq. ID No. 3, and the human intron non-coding OP-1 sequence shown in Seq. ID No. 1. The six human intron non-coding OP-1 sequences are at bases 3736 to 10700; bases 10897 to 35 11063; bases 11217 to 11424; bases 11623 to 13358; bases 13440 to 10548; bases 15166 to 17250; all of Seq. ID No. 1. In addition the portion of OP-1 sequence selected can be a combination of 5' non-coding, 3' non-coding and/or intron OP-1 sequence. Thus, the cell can have been transfected with a reporter gene in operative

association with a portion of 5' non-coding OP-1 genomic sequence that is independently selected from bases 3170 to 3317; 3020 to 3317; 2790 to 3317; 2548 to 3317; 2300 to 3317; 1300 to 3317; 1 to 3317; 2548 to 2790; 1549 to 2790; and 1 to 2790; all of Seq. ID 5 No. 1 or bases 2150 to 2296; 2000 to 2296; 1788 to 2296; 1549 to 2296; 800 to 2296; 1 to 2296; 1549 to 1788; 800 to 1788; 1 to 1788; all of Seq. ID No. 2. The lengths of bases within these 5' 10 non-coding sequences is selected to include portions of the sequence of DNA which was determined to be homologous between murine and human genomic OP-1, separately and as a part of a larger sequence including non-homologous DNA. Additionally, the portion of OP-1 sequence selected can be a portion of the region of homology between murine and human OP-1 DNA sequences, such as bases 2548-2790 or 2548-3317 of Seq. ID No. 1, or bases 1549 to 15 1788 or 1549 to 2296 of Seq. ID No. 2, and at least one of an Wt-1/Egr-1 consensus binding sequence, a TCC binding sequence, a FTZ binding sequence, a steroid binding sequence, and an intron. Thus the portion of OP-1 sequence selected can be a portion of the 5' 20 non-coding human or murine OP-1 genomic DNA sequences, as stated above, and at least one Wt-1/Egr-1 consensus binding sequence alone or in combination with at least one of a TCC binding sequence, a FTZ binding sequence, a steroid binding sequence, and a human OP-1 intron DNA sequence. In another embodiment more than one Wt-1/Egr-1 element is used, for example, between 1-6, or at 25 least six. These cells are suitable for use in the method of the invention.

In one embodiment, part of the OP-1 coding region is anticipated to have an expression regulatory function and also can be added to a vector for use in the screening assay described 30 herein. OP-1 protein is translated as a precursor polypeptide having an N-terminal signal peptide sequence (the "pre pro" region) which is typically less than about 30 amino acid residues, followed by a "pro" region which is about 260 amino acid residues, followed by the additional amino acid residues which comprise the 35 mature protein. The pre pro and pro regions are cleaved from the primary translation sequence to yield the mature protein sequence. The mature sequence comprises both a conserved C-terminal seven cysteine domain and an N-terminal sequence which varies significantly in sequence between the various morphogens. The

mature polypeptide chains dimerize and these dimers typically are stabilized by at least one interchain disulfide bond linking the two polypeptide chain subunits. After the pro domain is cleaved from the OP-1 protein it associates noncovalently with the mature 5 dimeric protein, presumably to enhance solubility and/or targeting properties of the mature species. See, for example, PCT/US93/07189, filed July 29, 1993. The pro region represents the nucleotide sequence occurring approximately 87 bases downstream of the ATG start codon, and continues for about 980 10 bases. The nucleotide sequence encoding the pro region is highly enriched in a "GC" sequence, which well may be competent to form a secondary structure (e.g., as part of the mRNA transcript) which itself may modulate OP-1 expression. Accordingly, part or all of the nucleotide sequence encoding an OP-1 pro region, particularly 15 that portion corresponding to a GC rich region, may be used, preferably in combination with one or more OP-1 non coding sequences, in the compositions and methods of the invention.

In another embodiment, the method can be practiced using a 20 cell known to express the OP-1 gene. Suitable DNA sequences for transfection are described below, as well as suitable cells containing transfected DNA sequences.

In another aspect, the invention provides molecules, 25 vectors, methods and kits useful in the design and/or identification of OP-1 expression modulating compounds. As used herein a "kit" comprises a cell transfected with a DNA sequence comprising a reporter gene in operative association with a portion of OP-1 upstream DNA sequence and the reagents necessary for detecting expression of the reporter gene. The portion of OP-1 upstream DNA chosen can be any of the various portions which have 30 been described herein.

Following this disclosure, medium flux screen assays, and 35 kits therefore, for identifying OP-1 expression modulating compounds are available. These compounds can be naturally occurring molecules, or they can be designed and biosynthetically created using a rational drug design and an established structure/function analysis methodology. The compounds can be amino acid-based or can be composed in part or whole of non-proteinaceous synthetic organic molecules.

The OP-1 expression modulating compounds thus identified then can be produced in reasonable quantities using standard recombinant expression or chemical synthesis technology well known and characterized in the art and/or as described herein. For 5 example, automated means for the chemical synthesis of nucleic and amino acid sequences are commercially available. Alternatively, promising candidates can be modified using standard biological or chemical methodologies to, for example, enhance the binding affinity of the compound for a DNA element and the preferred 10 candidate derivative then can be produced in quantity.

Once a candidate compound has been identified it can be tested for its effect on OP-1 expression. For example, a compound which upregulates (increases) the production of OP-1 in a kidney cell line is a candidate for systemic administration. The candidate 15 can be assayed in an animal model to determine the candidate molecule's efficacy *in vivo*. For example, the ability of a compound to upregulate levels of circulating OP-1 *in vivo* can be used to correct bone metabolism diseases such as osteoporosis (See, for example, PCT/US92/07932, supra). Useful *in vivo* animal 20 models for systemic administration are disclosed in the art and below.

As demonstrated herein below, OP-1 is differentially expressed in different cell types. Accordingly, it further is anticipated that a candidate compound will have utility as an inducer of OP-1 25 expression in one cell type but not in another. Thus, the invention further contemplates testing a candidate compound for its utility in modulating expression of OP-1 in different cells in vivo, including different cells known to express OP-1 under native physiological conditions.

30 Thus, in view of this disclosure, one of ordinary skill in recombinant DNA techniques can design and construct appropriate DNA vectors and transfect cells with appropriate DNA sequences for use in the method according to the invention to assay for compounds which modulate the expression of OP-1. These identified 35 compounds can be used to modulate OP-1 production and its available concentrations in both *in vivo* and *in vitro* contexts.

Brief Description of the Drawings

Fig. 1 shows the alignment of upstream sequences of the murine and human OP-1 gene. The murine sequence is present in the upper sequence lines and the human sequence is the lower sequence on all lines. The murine sequence is numbered backwards, counting back 5 from the first ATG of the translated sequence which is shown highlighted. For purposes of alignment, dashes are introduced into the DNA sequence, and three portions of human DNA sequence have been cut from the sequence and placed underneath a gap, below a solid triangle;

10 Fig. 2 shows a time course of murine uterus OP-1 mRNA regulation by estrogen; and

15 Fig. 3a shows a schematic of the 2 kb and 4 kb OP-1 mRNAs, the hybridization locations of probes 1 through 7 (indicated by the bars under the schematic). The solid line indicates OP-1 mRNA, the \* indicate potential poly A signals, the boxes indicate the translated portion of OP-1 mRNA with the hatched box showing the TGF- $\beta$  -like domain. The dashed lines indicate genomic DNA sequences. The arrows mark the locations of the cleavage site for OP-1 maturation.

20 Fig. 3b. shows a Northern blot hybridization analysis of OP-1 specific 2 kb and 4 kb mRNAs in murine uterine tissue. Lanes 1 through 7 correspond to probes 1 through 7 respectively. The 2 kb and 4 kb mRNAs are indicated by the 4- and 2-on the left side of Fig. 3b, and a 0.24 to 9.49 kb RNA size ladder is indicated by 25 dashes to the right of the figure.

Detailed Description

As will be more fully described below, we have identified regions in the OP1 genetic sequence useful in identifying 30 molecules capable of modulating OP-1 expression in vivo. Also as described herein, we have determined that OP-1 expression in vivo can be dependent both on cell type and on the status of the cell in a tissue. Specifically, as described herein below, OP-1 protein expression is differentially regulated in uterine tissue 35 depending on the status of the uterine tissue. For example, OP-1 expression is dramatically down-regulated in uterine mouse tissue during pregnancy, whereas it is normally expressed in this tissue in virgin mice. Moreover, OP-1 expression in other tissues such

as renal tissue apparently is unaffected during pregnancy. Administration of estrogen to a virgin mouse is capable of duplicating this down-regulation of OP-1 gene expression.

We investigated the DNA sequences responsible for the 5 regulation of OP-1 gene expression by cloning non-coding sequences for the human and mouse OP-1 gene. The tissue specific modulation of OP-1 gene expression, and the significant homology which was found between an approximately 750 base region of human and murine 5' non-coding OP-1 genomic sequence, implicate these sequences as 10 having utility in a method for the screening of compounds for their ability to modulate OP-1 expression.

In view of this disclosure and the examples provided below, a 15 method for identifying molecules which can affect OP-1 expression in a particular cell type in vivo now is provided.

#### Cloning of Human and Mouse OP-1 Gene Non-coding Sequences

In the Northern blot analysis of murine organs multiple OP-1 transcripts, are detected namely, three species of 1.8, 2.2, 2.4 20 kb and a prominent 4.0 kb RNA species (Özkaynak et al., 1992, J. Biol. Chem., 267:25220-25227; Özkaynak et al; Biochem. Biophys. Res. Comm., 179:116-123). The pattern is similar in rats with only the 1.8 kb species absent. The estrogen-mediated downregulation of OP-1 mRNA affects all of these species. In 25 order to prove that the 4.0 kb mRNA is in fact a transcript from the same OP-1 locus, cDNA clones were isolated from a mouse teratocarcinoma cDNA library.

Four independent clones were obtained that added sequence 30 information to the published mouse cDNA sequence. Two of these cDNA clones have longer 5'-untranslated sequences (0.4 and 0.3 kb) than previously reported (0.1 kb). Three of the murine clones contain additional 1.4 kb at the 3'-end. The combined sequences add up to a total OP-1 cDNA size of 3.5 kb, about 0.5 kb shorter than the 4.0 kb mRNA observed on Northern blots. cDNA clones that 35 represent the 2 kb and 4 kb messages are shown schematically in Figure 3a. Since the polyA-tail is lacking in those cDNA clones that extend the 3'-information, it was anticipated that missing 0.5 kb sequence occurs at the 3'-end.

In order to obtain the sequence immediately adjacent to the 3'-end of the 3.5 kb cDNA sequence, a mouse genomic library, ML1039J (Clontech), was screened with a 3'-end cDNA specific probe (0.45 kb, 3'-end Xmnl-EcoRI fragment of murine DP-1 cDNA) according to the parameters described below for the cloning of upstream non-coding sequences. This screen yielded four lambda clones which were analyzed by Southern blotting. All clones yielded a 1.5 kb Xmnl fragment which was subcloned from lambda Ö71 into a Bluescript vector and sequenced. Three polyadenylation signals (AATAAA) (Proudfoot et al, (1976) Nature, 263:211-214) were found in this genomic fragment, at 3.52-, 3.58-, and 3.59 kb (shown schematically in Fig. 3a by the \*). The 3'-end cDNA and the genomic DNA sequences in the 1.5 kb Xmnl fragment overlap by 0.4 kb in a region that immediately precedes the second polyadenylation signal located at 3.5 kb (Figure 3a, region indicated by probe 6) and are in complete agreement within this stretch.

Human upstream non-coding sequence and additional mouse upstream non-coding sequence were obtained by screening human and mouse genomic libraries, HL1067J and ML1030J respectively (Clontech). All libraries were screened by an initial plating of 750,000 plaques (approximately 50,000 plaques/plate). Hybridizations were done in 40% formamide, 5 x SSPE, 5 x Denhardt's solution, and 0.1% SDS at 37°C. Nonspecific counts were removed in 0.1 x SSPE, 0.1 % SDS by shaking at 50°C. Human and mouse upstream genomic DNA sequences were obtained from clones lambda Ö3 and lambda Ö33, respectively (Clontech, HL1067J and ML1030J). These lambda clones were isolated using a <sup>32</sup>P-labeled probe made from a human 0.47 kb EcoRI OP-1 cDNA fragment (obtained from pÖ115) containing mainly 5' non-coding and exon 1 sequences.

A 7 kb EcoRI fragment from the human genomic clone, lambda Ö3, was isolated which contains 5 kb of upstream non-coding sequence. Additional upstream sequence information for murine was obtained by subcloning a 1.1 kb PstI fragment from the genomic phage clone lambda Ö33. This fragment overlaps with the 5'-end of the longest murine cDNA clone by 0.3 kb in the 5' non-coding region and provided 0.8 kb additional sequence information. A schematic diagram of the 2- and 4 kb OP-1 messages is shown in Figure 3a

with dashed lines indicating supplementing information derived from murine upstream and downstream genomic DNA.

All sequencing was done according to Sanger et al. (1977) Proc. Natl. Acad. Sci. 74:5463-5467, using exonuclease III-  
5 mediated unidirectional deletion (Özkaynak et al., (1987) BioTechniques, 5:770-773), subcloning of restriction fragments, and synthetic primers. Compressions were resolved by performing the reactions at 70°C with Taq polymerase and using 7-deaza-GTP (U.S. Biochemical Corp., Cleveland, OH).

10

Verification of OP-1 mRNA Sequences by Northern Blotting

To verify the structures of the short and long mRNA species observed, Northern blot hybridizations were performed with probes made from seven non-overlapping DNA fragments (Fig. 3a; probes 1  
15 through 7) specific to the 5' and 3' non-coding region, the protein coding sequence, and genomic regions upstream or downstream of the predicted mRNAs, respectively.

Hybridization of these probes to individual Northern blot strips containing mouse kidney mRNA is consistent with the  
20 predicted 4 kb mRNA structure. As shown in Fig. 3a, and Fig. 3b, the genomic DNA probes 1 and 2 did not hybridize to any message. Probe 2 is specific to the upstream sequences immediately adjacent to the cDNA. Probes 3, 4, and 5, specific to 5' non-coding, coding, and 3' non-coding regions, respectively, hybridized to  
25 both the 2 kb and 4 kb messages, hence these sequences are present in both messages. Probe 6, specific to sequences between the first and second polyadenylation signals, hybridized only to the 4 kb message. Finally, probe 7 which is specific to sequences further downstream of the fourth (last) polyadenylation signal,  
30 did not hybridize to any message. The results obtained with these probes confirm the two OP-1 mRNA structures and the approximate 5'- and 3'-end boundaries of OP-1 transcripts shown in Figure 3a. This demonstrates that the 2 kb and 4 kb mRNA's are from the same OP-1 genomic locus rather than from multiple genes.

35 The extensive 3' sequence included in the 4 kb mRNA transcript suggests that the 3' untranslated sequence may play a role in OP-1 gene expression particularly as it has been detected across species namely, in mouse, rat, dog, human and chicken. Multiple

stop codons in all three possible translation reading frames rule out the likelihood that this sequence encodes a peptide. The untranslated sequence itself may act therefore to influence mRNA stability.. For example, the sequence may interact with another 5 protein as has been described for transferrin receptor mRNA. Here, IRE-binding protein (IRE; iron response element) stabilizes the transferrin receptor mRNA by binding to the 3'-end of the mRNA (Standard et al., 1990, Genes Dev., 4:2157-2168, incorporated herein by reference). Alternatively, the 3'-end sequences may be 10 interacting with the 5'-end sequences thereby affecting initiation of protein synthesis or, the 3'-end sequences may be serving as a binding site for other RNAs which can interfere with the binding of an expression in modulating molecule, including repressor molecule. (Klausner et al., 1989, Science, 246:870-872; Kozak, 15 1992, Ann. Rev. Cell Biol., 8:197-225, incorporated herein by reference).

Comparison of 5' Non-coding Sequences of Human and Mouse OP-1 DNA

The cloning of the 5' non-coding genomic murine and human OP-1 20 DNA sequences demonstrated that a high degree of sequence homology exists between the human and murine 5' non-coding DNA sequences. The homology extends from the base immediately upstream of the translation start site for the OP-1 morphogen protein to approximately 750 bases upstream of the translation start site, as 25 is shown in the shaded regions of Fig. 1, with the murine sequences being the upper lines and the human sequences being the lower lines. The 5' nucleotide of the region of homology for the human OP-1 5' non-coding sequence is base 2548 of Seq. ID No. 1 and for the murine OP-1 5' non-coding sequence is base 1549 of 30 seq. ID No. 2. The significant homology between the human and murine 5' non-coding sequences of OP-1 suggest that this region may be important in the regulation of OP-1 expression. As will be discussed in more detail below, this region contains several 35 conserved DNA sequences which have been identified as the DNA binding sequences for two DNA binding proteins, Wt-1 and Egr-1, which both recognize these DNA sequences. The DNA binding sequences for Wt-1/Egr-1 present in human and murine are marked in Fig. 1 with a single line. Also, the TCC binding sequence, a DNA binding sequence for Wt-1 and Egr-1, is marked in Fig. 1 by the

double line. WT-1 and Egr-1 proteins have also been implicated in the regulation of expression of several genes which are unrelated to OP-1.

Alignments of mouse and OP-1 human genomic sequences reveals a 5 conserved stretch of 0.75 kb just upstream of the first ATG that contains several patterns with marked similarity to the zinc-finger protein binding sequence (5'-GCG GGG GCG-3') specific for Egr-1 and Wt-1 (Christy et al., 1989, PNAS, 86:8737-8741; Rauscher et al., 1990, Science, 250:1259-1262; Drummond et al., 1992, Science, 257:664-678). In mouse, a total of 8, and in human 7, 10 patterns, conforming to the degenerate Egr-1/Wt-1 binding sequence (5'-GNG NGG GNG-3') (Rupprecht et al., 1994, J. Biol. Chem., 269: 6198-6202; Werner et al., 1994, J. Biol. Chem., 269: 12940-12946 are located before and after the presumed transcriptional 15 initiation site (Fig. 1, shown by solid single lines). The presence of these has significance in light of the elevated levels of Wt-1 mRNA in the rat uterus decidua during pregnancy (Zhou et al., 1993, Differentiation, 54:109-114).

The analysis also revealed, in the human upstream region, a 20 pattern of seven TCC repeats, present at -561, immediately 3' of two Egr/Wt-1 sequences (at -624 and -587) (Figure 1, shown by double solid lines and at position 2758-2778 of Seq. ID No. 1). The mouse upstream region contains a similar, albeit less obvious sequence at -356 and at position 1755-1769 of Seq. ID No. 2. This 25 TCC-repeat pattern is found in the promoters of PDGF-A and several other growth-related genes, and Wt-1 has been found to activate transcription when either of the sequences are present and to suppress it when both sequences are present. (Wang et al., 1992, Biochem. Biophys. Res. Comm., 188:433-439; Wang et al. 1993, PNAS, 90:8896-8900 incorporated herein by reference). Accordingly, 30 estrogen receptor may exert its effect on OP-1 expression in uterus by upregulating Wt-1, either directly or indirectly. Alternatively or, in addition other regulatory elements, located 35 further upstream of the OP-1 gene may be involved in estrogen regulation.

Also on Fig. 1, the human 5' non-coding DNA sequence is shown to contain a Fushi-tarazu (FTZ) binding sequence which is marked by carats below the human DNA sequence. A FTZ binding sequence is

bound by the Fushi-tarazu protein (FTZ-F1), which is a member of the superfamily of nuclear receptors (Parker, (1993) Current opinion in Cell Biology, 5:499-504, ). The superfamily of nuclear receptor proteins include steroid hormones, retinoids, thyroid 5 hormone, nerve growth factor and Fushi-tarazu, and are structurally related. FTZ-F1 is likely to belong to a subfamily of nuclear receptors that bind DNA as monomers.

The FTZ-F1 protein is a positive regulator at the fushi-tarazu gene in blastoderm stage embryos of *Drosophila*. FTZ-F1 is 10 closely related in the silkworm (*Bombyx*) BmFTZ-F1 protein and the mouse embryonal long terminal repeat binding protein (ELP) and all of them are members of the nuclear hormone receptor superfamily, which recognizes the same 9 base pair sequence, 5'-PyCAAGGPyCPu- 15 3'. The FTZ binding sequence does not apparently have a direct or inverted repeat. In contrast, other members of the nuclear hormone receptor superfamily usually bind to repeated sequences. Nevertheless, the FTZ-F1, BmFTZ-F1 and ELP proteins have high affinities for the FTZ binding site DNA, indicating that the mechanism that the binding is somewhat different from that of 20 other members of the nuclear hormone receptor superfamily. (Hitachi et al., 1992, Mol. and Cell Biology December, pp. 5667-5672.).

The mRNA transcription initiation site for human OP-1 is marked on Fig. 1 by the upward arrow, and the OP-1 protein 25 translation initiation site is marked on Fig. 1 by the solid triangles just prior to the highlighted ATG. The transcription initiation site for the human OP-1 gene is at base 2790 of Seq. ID No. 1 and the analogous site for murine is at base 1788 of Seq. ID No. 2. The translation initiation site for the human OP-1 gene is 30 at base 3318 of Seq. ID No. 1 and for murine it is at base 2296 of Seq. ID No. 2. The high degree of identity that the murine and human DNA sequences share in the region between the transcription initiation site and the translation initiation site, suggests that 35 this region likely plays a role in the modulation of the expression of the OP-1 gene product.

Analysis of OP-1 Gene Expression in Mouse Tissues

A detailed analysis of the uro-genital tract of rats has revealed OP-1 mRNA expression in the renal (kidney), and bladder tissues, as well as at other sites of the urogenital organ system. The most abundant levels are present in renal and uterine tissue 5 (8 week old mice), while much lower levels were found in ovaries. The mRNA level of G3DPH, a "housekeeping function" molecule, was used as an internal control for recovery and quality of mRNA preparations and equal amounts of poly(A)+ RNA (5mg), were loaded into each lane.

10 Preparation of RNA and Northern blot hybridization analysis was conducted as follows. 8-week-old female mice, strain CD-1, were obtained from Charles River Laboratories, Wilmington, MA. Total RNA, from the various organs of mice was prepared using the acid-guanidine thiocyanate-phenol-chloroform method (Chomczynski 15 et al., (1987) *Anal. Biochem.* 162:156-159). The RNA was dissolved in TES buffer (10 mM Tris-HCl, 1 mM Na<sub>2</sub>-EDTA, 0.1% SDS, pH7.5) containing Proteinase K (Stratagene, La Jolla, CA; approx. 1 mg proteinase /ml TES) and incubated at 37°C for 1 hr. Poly (A)<sup>+</sup> RNA was selected in a batch procedure on oligo(dT)-cellulose 20 (Stratagene, La Jolla, CA) in 0.5 M NaCl, 10 mM Tris-HCl, 1 mM Na<sub>2</sub>-EDTA, pH 7.4 (1 x binding buffer). For the selection of poly (A)+ RNA, total RNA obtained from 1 g of tissue was mixed with approximately 0.1g of oligo(dT)-cellulose (in 11 ml TES containing 0.5 M NaCl). The tubes containing the RNA and oligo(dT)-cellulose 25 were gently shaken for approx. 2 hrs. Thereafter, the oligo(dT)-cellulose was washed twice in 1x binding buffer and once in 0.5x binding buffer (0.25 M NaCl, 10 mM Tris-HCl, 1 mM Na<sub>2</sub>-EDTA, pH 7.4) and poly (A)+ RNA was eluted with water and precipitated with ethanol.

30 Poly(A)+ RNA (5 mg per lane) was electrophoresed on 1.2% agarose-formaldehyde gels with 1 mg of 400 µg/ml ethidium bromide added to each sample prior to heat denaturation (Rosen et al., (1990) *Focus*, 12:23-24). Electrophoresis was performed at 100 Volts with continuous circulation of the 1 x MOPS buffer (Ausubel 35 et al., eds., (1990) *Current Protocols in Molecular Biology*, John Wiley & Sons, New York). Following electrophoresis, the gels were photographed, rinsed briefly in water, and blotted overnight onto Nytran (Schleicher & Schuell Inc., Keene, NH) or Duralon-UV

(Stratagene) membranes in 10 x SSC. The membranes were dried at 80° for 30 min. and irradiated with UV light (1 mW<sup>2</sup>/cm<sup>2</sup> for 25 sec.).

The <sup>32</sup>P-labeled probe was made from a murine OP-1 cDNA fragment (0.68 kb BstXI-BG1I frg.) by random hexanucleotide priming (Feinberg et al., (1984) Anal. Biochem., 137:266-267). The hybridizations were done in 40% formamide, 5x SSPE, 5x Denhardt's, 0.1% SDS, pH 7.5 at 37°C overnight. The non-specific counts were washed off by shaking in 0.1x SSPE, 0.1% SDS at 50°C. For re-use, filters were stripped in 1 mM Tris-HCl, 1 mM Na<sub>2</sub>-EDTA, 0.1% SDS, pH 7.5 at 80° C for 10 min.

#### Analysis of OP-1 Expression During Pregnancy in Mice

An examination of the effect of pregnancy upon OP-1 expression was undertaken by measuring OP-1 mRNA levels in kidney, ovary and uterus, before, during, and after pregnancy (virgins, 2-day post-coital (pc), 4-day pc, 6-day pc, 8-day pc, 13 day pc, 17-day pc, 3-day lactating, and retired breeders) by Northern blot hybridization of poly(A)+ RNA. These measurements demonstrated that, while kidneys show no pregnancy-related changes in OP-1 mRNA levels, the uterine levels became nearly undetectable by 6-day pc. However, no changes were observed in the ovaries. A dramatic and rapid decline in OP-1 message in uterine tissue between day 3 and 4 of pregnancy is apparent in the comparison with virgin animals.

The levels of OP-1 mRNA in the embryo and maternal levels in uterus of 8 week old mice at day 13 and 16 of the pregnancy were also compared. While the OP-1 expression in the pregnant uterus is dramatically reduced, high levels of OP-1 message are found in the mouse embryo at 13- and 16-days. Thus, at a stage of pregnancy when OP-1 mRNA expression in the maternal uterus is almost undetectable, embryonal OP-1 expression is high. The high embryonal OP-1 expression also is detected consistent with the relatively high levels of OP-1 mRNA, found in human placenta. The level of OP-1 mRNA measured in the embryo is in the same range as that measured in adult kidney or virgin uterus tissue. Hence, it is likely that OP-1 plays a critical role in the development of the embryo which may require appropriate amounts of OP-1 at very specific stages of tissue and organ morphogenesis. While not being

limited to any given theory, it is possible that OP-1 expression in uterine tissue during pregnancy potentially could interfere with the level of OP-1 produced by the developing embryo, and thereby interfere with proper development of the embryo.

5 Therefore, a shut-down or inhibition of uterine OP-1 expression during pregnancy might be for the benefit of the fetus.

Effect of Estrogen and Progesterone on OP-1 Expression

10 During pregnancy the estrogen and progesterone levels increase many fold and high levels are sustained until birth. To determine whether these hormonal changes are responsible for the altered OP-1 transcription in pregnant uterine tissue, non-pregnant female mice were subcutaneously administered 17 $\beta$ -estradiol, or progesterone, or a combination of both.

15 In the first experiment the rapid increase in estrogen and progesterone levels during pregnancy was simulated. Non-pregnant mice were injected subcutaneously on four consecutive days with increasing doses, starting with 20 mg 17 $\beta$ -estradiol, or 100 mg progesterone or the combination of both and doubling the dose on 20 each following day. On the fourth day the animals were sacrificed and mRNA was isolated from uteri and kidneys. A striking negative effect of 17 $\beta$ -estradiol on the uterine OP-1 mRNA expression was observed, but no effect by progesterone was seen. In the kidneys, however, mRNA levels did not change after 17 $\beta$ -estradiol or 25 progesterone treatment.

Another experiment addressed the time course: 17 $\beta$ -estradiol was administered to virgin female mice at a constant dose of 200 mg (50 ml of 4 mg/ml 17 $\beta$ -estradiol per day, subcutaneously in DMSO [dimethyl sulfoxide] + 150 ml 150 mM NaCl) (Figure 2). Following 30 this, their uteri were extracted, poly(A)+ RNA was prepared, equal amounts of poly(A)+ RNA (5 mg) was loaded into each lane of a 1.2% agarose-formaldehyde gel and analyzed by Northern blot hybridization. The effect was rapid, with considerable decrease of OP-1 mRNA 12 hours after administration of 17 $\beta$ -estradiol and 35 almost undetectable levels by 48 hours, as shown in Fig. 2. In the figure, the lanes correspond as follows: from left to right, 0-day (negative control), 0-day (negative control), 0.5-, 1-, 2-,

3-, 4-, 5-, 6-, 7-, and 8-days. The arrowheads mark the two major OP-1 mRNA species. A modest amount of message reappears a few days later (Figure 2).

The uterus has been identified as a major site of OP-1 expression. The level of OP-1 expression in uterine tissue is comparable to that observed in renal tissue. However, during pregnancy, by day four, the uterine OP-1 mRNA levels are reduced to the limit of detection. The loss of OP-1 expression corresponds with also is rising levels of estrogen during this same time frame. The same dramatic loss of uterine OP-1 message also is observed in estrogen-treated animals, suggesting that estrogen is involved in negative regulation of OP-1 expression in uterine tissue. The effect of estrogen is rapid, with most of the message disappearing after 12 hours of 17 $\beta$ -estradiol administration. The reappearance of some OP-1 message at later days may be due to a counter-regulatory mechanism. In contrast to the modulated OP-1 mRNA levels in the uterus, no substantial changes occur in renal tissue during pregnancy or in response to estrogen treatment. Therefore, OP-1 mRNA expression in these different organs is regulated independently. The differential expression may be due, for example, to a lack of estrogen receptors in renal tissue. Alternatively, co-regulation by means of one or more accessory molecules that interact with estrogen or a related nuclear receptor molecule(s) may allow for the independent regulation. For example, each of Wt-1 protein (which binds to the Wt-1/Egr-1 element) and OP-1 protein are required for normal kidney development, and each are expressed at high levels during kidney tissue development. As described above the OP-1 promoter region contains Wt-1 consensus binding elements. Wt-1 protein also has been shown to negatively regulate the transcription of the insulin growth factor II gene and the platelet-derived growth factor A chain gene. Kreidberg et al., Cell, 1993, 74:679-691. Without being limited to a given theory, it may be that Wt-1 protein, either alone or in combination with one or more molecules is involved in the expression of OP-1. For example, Wt-1 protein may act in concert with a nuclear hormone receptor element, including, for example, the estrogen receptor element.

Implications of Tissue Specific Differential Regulation of OP-1 Expression

Estrogen also has been shown to inhibit the uterine expression of calbindin-D<sub>28k</sub>, a vitamin D dependent calcium binding protein, 5 the  $\alpha$ -subunit expression of the glycoprotein hormones, and other proteins involved in bone formation. Estrogen also has been shown to cause dramatic decreases in the steady state mRNA levels of the bone matrix proteins osteocalcin, prepro  $\alpha$ 2(I) chain type I collagen, osteonectin, osteopontin, and alkaline phosphatase in an 10 ovariectomized rat, which is a rat model for osteoporosis.

Estrogen appears to mediate its beneficial effect on bone metabolism in the osteoporotic model through inhibition of osteoclasts. Estrogen does not reverse osteoporosis. By contrast, OP-1, which is expressed in uterine, renal and bone 15 tissues, is able to induce an increase in bone mass in the osteoporotic model. Thus, the negative effect of estrogen on OP-1 expression in uterine tissue may seem unexpected in view of estrogen's effect on bone metabolism.

In addition to the 5' non-coding DNA sequences of OP-1, the 20 other non-coding sequences such as introns and 3' non-coding sequences may be involved in the modulation of OP-1 protein expression. This invention presents a method in which these non-coding sequences are assayed while in operative association with a reporter gene for their influence on the expression of OP-1. Non- 25 coding sequences which are involved in the modulation of OP-1 expression will be identified by culturing cells transfected with the non-coding sequences, in operative association with a reporter gene, with one or more compound(s), measuring the level of reporter gene expression, and comparing this level of expression 30 to the level of reporter gene expression in the absence of the compound(s).

EXEMPLARY CELLS, VECTORS, REPORTER GENES AND ASSAYS FOR USE IN  
SCREENING COMPOUNDS WHICH MODULATE OP-1 REGULATORY SEQUENCES  
35

I. Useful Cells

Any eukaryotic cell, including an immortalized cell line suitable for long term culturing conditions is contemplated to be useful for the method and cell of the invention. Useful cells

should be easy to transfect, are capable of stably maintaining foreign DNA with an unarranged sequence, and have the necessary cellular components for efficient transcription and translation of the protein, including any elements required for post- 5 translational modification and secretion, if necessary. Where the cell is to be transfected with a non-dominating selection gene, the cell genotype preferably is deficient for the endogenous selection gene. Preferably, the cell line also has simple media composition requirements, and rapid generation times.

10 Particularly useful cell lines are mammalian cell lines, including myeloma, HeLa, fibroblast, embryonic and various tissue cell lines, e.g., kidney, liver, lung and the like. A large number of cell lines now are available through the American Type Culture Collection (Rockville, MD) or through the European Collection of 15 Animal Cell Cultures (Porton Down, Salisbury, SP4 0JG, U.K.)

Where, as here, the expression of a reporter gene that is controlled by non-coding sequences of the morphogen OP-1 is to be analyzed, particularly useful cells and cell lines are envisioned to include eukaryotic, preferably mammalian cells of a tissue and 20 cell type known to express OP-1 and/or closely related proteins. Such cells, include, without limitation, cells of uro-genital cell origin, including kidney, bladder and ovary cells, lung, liver, mammary gland and cardiac cells, cells of gonadal origin, cells of gastrointestinal origin, glial cells and other cell lines known to 25 express endogenous genes encoding morphogenic proteins. Preferred cell lines are of epithelial origin.

## II. Exemplary Vectors/Vector Construction Considerations

Useful vectors for use in the invention include, but are not 30 limited to cosmids, phagemids, yeast artificial chromosomes or other large vectors. Vectors that can be maintained within the nucleus or integrated into the genome by homologous recombination are also useful. For example a vector such as PSV2CAT would be useful.

35 Selected portions of non-coding OP-1 sequence can be cloned into a useful vector using standard molecular cloning techniques, as will be apparent to one of ordinary skill in the art. Restriction endonuclease sites will be utilized when possible, and can be engineered into the sequence when needed. If restriction

endonuclease sites are needed to be engineered into the sequence, eight base recognition sites are preferable because they generally occur infrequently in DNA and will enhance a practitioners ability to obtain the sequence of interest. Restriction endonuclease 5 sites can be engineered into the non-coding sequence using the common techniques such as site directed mutagenesis and PCR with primers including the desired restriction endonuclease site.

As discussed above, murine and human OP-1 sequences share a region of high homology covering approximately 750 bases upstream 10 of the translation initiation site as shown by the shading in Fig. 1. This region is positions 2548-3317 of Seq. ID No. 1 and positions 1549-2296 of Seq. ID No. 2. The mRNA transcription initiation site lies within this region at position 2790 of Seq. ID No. 1 and by analogy at position 1788 of Seq. ID No. 2, shown 15 in Fig. 1 by the upward arrow. This suggests that positions 2548-2790 of Seq. ID No. 1 and 1549-1788 of Seq. ID No. 2 contain conserved promoter elements for the expression of OP-1 mRNA, and approximately 500 bases at positions 2791-3317 of Seq. ID No. 1 and positions 1790-2296 of Seq. ID No. 2 contain conserved 20 elements of the transcribed, but not translated, sequences all or part of which may be involved in the regulation of OP-1 expression. Additionally sequences upstream of the homology region may also be involved in the regulation of OP-1 expression. Thus a range of upstream sequences, including sequences upstream 25 of the transcription initiation site and not including the approximately 500 bases of transcribed sequence, can be fused in operative association with a reporter gene to modulate expression of the gene.

3' non-coding sequences and intron sequences also can be fused 30 in operative association with a reporter gene, either separately or in combination with each other or with 5' non-coding sequences. For example, one can place the 5' sequences defined by positions 2790-3317; 2548-2790 or 2548-3317 of Seq. ID No. 1, and either/both of 3' sequences or intron sequences in operative 35 association with a reporter gene. The positions of the six introns are shown in Seq. ID No. 1 as bases 3736 to 10700; bases 10897 to 11063; bases 11217 to 11424; bases 11623 to 13358; bases 13440 to 10548; bases 15166 to 17250;

Also envisioned is a nucleic acid construct comprising a small fragment of 5' non-coding OP-1 sequence in combination with additional conserved elements such as one or more Wt-1/Egr-1 binding sequences; a TCC binding sequence and/or a FTZ binding 5 sequence in operative association with a reporter gene. Such a nucleic acid construct also could include intron sequences and/or 3' non-coding sequences.

A range of useful 5' non-coding fragments has been provided, and as will be apparent to those of ordinary skill in the art, 10 smaller fragments of OP-1 sequence also are useful. Such smaller fragments can be identified by deleting bases from one or both ends of the provided 5' non-coding fragments, using techniques that are well known in the art and testing the truncated constructs for their ability to modulate reporter gene expression. 15 In this way, the shortest modulating sequences can be identified.

### III. Transfection Considerations

Any method for incorporating nucleic acids into cells of interest is contemplated in the method of the invention. Calcium 20 phosphate (CaPO<sub>4</sub>), followed by glycerol shock is a standard means used in the art for introducing vectors, particularly plasmid DNA into mammalian cells. A representative method is disclosed in Cockett et al., (1990) Biotechnology 8: 662-667, incorporated herein by reference. Other methods that may be used include 25 electroporation, protoplast fusion, particularly useful in myeloma transfections, microinjections, lipofections and DEAE-dextran mediated uptake. Methods for these procedures are described in F.M. Ausubel, ed., Current Protocols in Molecular Biology, John Wiley & Sons, New York (1989).

30 As will be appreciated by those having skill in the art, optimal DNA concentrations per transfection will vary according to the transfection protocol. For calcium phosphate transfection, for example, preferably 5-10 µg plasmid DNA per plasmid type is transfected. In addition, the DNA to be transfected preferably is 35 essentially free of contaminants that may interfere with DNA incorporation. A standard means used in the art for purifying DNA is by ethidium bromide banding.

## IV. Exemplary Reporter Genes

There are numerous reporter systems commercially available, which include, without limitation, the chloramphenicol acetyltransferase (CAT), luciferase, GAL4, and the human growth

5 hormone (hGH) assay systems.

CAT is a well characterized and frequently used reporter system and a major advantage of this system is that it is an extensively validated and widely accepted measure of promoter activity. See, for example, Gorman, C.M., Moffat, L.F., and

10 Howard, B.H. (1982) Mol. Cell. Biol., 2:1044-1051 for a description of the reporter gene and general methodology. In this system cells are harvested 2-3 days after transfection with CAT expression vectors and extracts prepared. The extracts are incubated with acetyl CoA and radioactive chloramphenicol.

15 Following the incubation acetylated chloramphenicol is separated from nonacetylated form by thin layer chromatography. In this assay the degree of acetylation reflects the CAT gene activity with the particular promoter.

Another well-recognized reporter system is the firefly 20 luciferase reporter system. See, for example Gould, S.J., and Subramani, S. (1988) Anal. Biochem., 7:404-408 for a description of the reporter gene and general methodology. The luciferase assay is fast and has increased sensitivity. The system also is particularly useful in bulk transfections or if the promoter of 25 interest is weak. In this assay transfected cells are grown under standard conditions, and when cultured under assay conditions both ATP and the substrate luciferin is added to the cell lysate. The enzyme luciferase catalyzes a rapid, ATP dependent oxidation of the substrate which then emits light. The total light output is 30 measured using a luminometer according to manufacturer's instructions (e.g., Cromeega) and is proportional to the amount of luciferase present over a wide range of enzyme concentrations.

A third reporter system is based on immunologic detection of 35 hGH, it is quick and easy to use. (Selden, R., Burke-Howie, K. Rowe, M.E., Goodman, H.M., and Moore, D.D. (1986), Mol. Cell. Biol., 6:3173-3179 incorporated herein by reference). hGH is assayed in the media, rather than in cell extracts. This allows

direct monitoring over by a single population of transfected cells over time.

As indicated above and as will be appreciated by those having ordinary skill in the art, particular details of the conventional 5 means for transfection, expression, and assay of recombinant genes are well documented in the art and are understood by those having ordinary skill in the art. The instant invention enables and discloses vectors, cells and a method for screening compounds to determine the capability of compounds to modulate the expression 10 of OP-1 via the non-coding sequences of the OP-1 genomic DNA.

Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in mammalian expression systems can be found in a number of texts and laboratory manuals in the art, such as, for example, F.M. Ausubel 15 et al., Ed., Current Protocols in Molecular Biology, John Wiley & Sons, New York, (1989).

#### VIII. Exemplary Homologous/Non-Homologous Recombination

One approach to screen for inducers of (organ-specific) OP-1 20 expression in a particular cell line derived from a particular tissue such as renal or uterine tissue, is through gene targeting by homologous recombination (Sedivy et al., W.H. Freeman & Co., New York (1992); A.S. Waldman, Crit. Rev. Oncol. Hematol. 12, 49 (1992)). In one strategy the endogenous (genomic) OP-1 gene is 25 replaced by another reporter gene which is optimally suited for screening assays, such as the firefly luciferase gene. To target the OP-1 gene in an appropriate cell line, e.g., a kidney cell line or NBT-2, the following arrangement of genetic elements can be assembled.

30 Genomic OP-1 upstream and promoter sequences preferably 3000 to 5000 nucleotides in length, and which mediate the homologous recombination, are attached to the luciferase gene. The OP-1 upstream sequences down to the first coding ATG can be attached at the start codon ATG of the luciferase coding sequence, using a 35 restriction site such as Ncol, which can be introduced by site directed mutagenesis into both the promoter and the luciferase sequences.

Also included is a selective marker, preferably the neo gene, without its own promoter. Preferably, selectable marker (neo) is placed downstream of the reporter gene (luciferase), after an intercistronic sequence derived from the poliovirus genome and 5 which allows translation of the sequence marker on the same transcript as the reporter gene transcripts. Details of this approach, including specific intercistronic sequences and the detailed steps of homologous recombination, are described in the art, including (Jasin et al., PNAS USA 85:8583 (1988); Sedivy et 10 al., PNAS USA 86, 227 (1989); Dorin et al., Science 243:1357 (1989) the disclosures of which are incorporated herein by reference. As described therein, the endogenes OP-1 gene is replaced by the luciferase and neo coding sequences and the expression of these sequences then assayed in a standard A 15 screening protocol.

A genetic arrangement of OP-1 promoter (as much genomic OP-1 upstream sequence as possible, up to 10,000 bp) and reporter gene (without its original promoter but joined directly to the OP-1 ATG or in its vicinity) can also be introduced into cells on standard 20 eukaryotic expression vectors. These vectors carry selectable markers (neo, dhfr, etc.) and will typically be integrated into the host genome with variable copy number ranging from one to several copies without efforts at amplification. Also, if desired, the vector or gene copy number can be enhanced using a 25 well characterized amplifiable gene, such as dhfr in conjunction with methotrexate. Commercial vectors designed for autonomous replication without integration are readily available. One source vector is the Episomal Expression Epstein Barr Virus Vector (pREP, Invitrogen Corp., San Diego CA).

30 Introns also can be tested for regulatory sequences as described hereinabove using the methods described herein. One or more intron sequences derived from a genomic OP-1 locus preferably is introduced into proper mammalian cells using, for example, a yeast artificial chromosome (pYACneo, Clontech, Inc. Palo Alto, 35 CA) (Ref. Albertson, H.M. et al. PNAS USA, 87:4256, 1990), or other vectors adapted to allow transfer of large sequences, e.g., up to 1 megabases. As for the OP-1 5' or 3' noncoding sequences described above, the intron sequence or a portion thereof is incorporated in operative association with a reporter gene and the

ability of the sequence to modulate reporter gene expressions then associated.

5       X. Exemplary Screening Assay for Compounds which Alter OP-1 Gene or Reporter Gene Levels

Candidate compound(s) which may be administered to affect the level of a given endogenous morphogen, such as OP-1, or a reporter gene that is fused to OP-1 non-coding sequence may be found using the following screening assay, in which the level of reporter gene 10 production by a cell type which produces measurable levels of the reporter gene expression product by incubating the cell in culture with and without the candidate compound, in order to assess the effects of the compound on the cell. This can be accomplished by detection of the reporter expression product either at the protein 15 or RNA level. The protocol is based on a procedure for identifying compounds which alter endogenous levels of morphogen expression, a detailed description also may be found in PCT US 92/07359.

20       Cultured cells are transfected with portions of OP-1 non-coding sequences in operative association with a reporter gene, and such transfected cells are maintained with the vector remaining as a plasmid in the cell nucleus or the vector can be integrated into the host cell genome, preferably at the OP-1 genomic locus.

25       Cell samples for testing the level of reporter gene expression are collected periodically and evaluated for reporter gene expression using the appropriate assay for the given reporter gene as indicated in the section describing reporter gene assays, or, alternatively, a portion of the cell culture itself can be 30 collected periodically and used to prepare polyA(+) RNA for mRNA analysis.

Once candidate compounds are identified, they can be produced in reasonable, useful quantities using standard methodologies known in the art. Amino acid-based molecules can be 35 encoded by synthetic nucleic acid molecules, and expressed in a recombinant expression system as described herein above or in the art. Alternatively, such molecules can be chemically synthesized, e.g., by means of an automated peptide synthesizer, for example.

Non-amino acid-based molecules can be produced by standard organic chemical synthesis procedures.

Provided below is an exemplary protocol for carrying out the method of the invention, using the CAT gene as the reporter gene 5 and one or more mammalian cell lines known to express OP-1. The example is non limiting, and other cells, reporter genes and OP-1 non-coding sequences are envisioned.

#### Exemplary Construction Of Representative Vectors For Transfections

A DNA fragment containing the OP-1 promoter can be joined to a 10 reporter gene for transfection into a cell line that expresses endogenous OP-1. Suitable cell lines are selected by Northern blot hybridization to an OP-1 specific probe (by analyzing the cell extracts for OP-1 mRNA). Using this technology we have found several cell lines which make high levels of OP-1 mRNA, and some 15 of these lines are the kidney line IMCD, the bladder line NBT II.

An approximately 5 Kb EcoRI, BamHI genomic fragment containing approximately 4 Kb of upstream OP-1 sequences as well as part of the first intron is blunt-ended with T4 DNA polymerase and cloned into a polylinker of a pUC vector (pO146-1). An approximately 3.5 20 kb DNA fragment containing human OP-1 upstream sequences is obtained by deleting a portion of coding sequences and the first intron from pO146-1 with the restriction enzyme EheI. The ~3.5kb fragment has blunt ends and contains mostly 5' non-coding sequences and also includes a short stretch of 30 bases into the 25 OP-1 gene. This upstream fragment is of ~3.5kb ligated to a 1.6 kb HindIII-BamHI fragment from the CAT gene obtained from the vector SV2CAT by 5' HindIII end blunted ligation. The 1.6kb CAT gene fragment contains about 70 bases of upstream sequences. These ligated fragments are cloned into Bluescript KS(-) vector 30 (Stratgene, La Jolla, CA). This construct in turn is subjected to site specific mutagenesis to delete the extra sequences (approximately 30 bases) from the 3' end of the OP-1 upstream sequences and the adjacent 5' non-coding sequences (approximately 70 bases) from the CAT gene. This mutagenesis results in the 35 elimination of any OP-1 coding sequences from the promoter fragment as well as any non-coding sequences upstream of the CAT gene. Thus the resulting construct is a fusion of OP-1 upstream sequences with the CAT gene sequences which encode the CAT

protein. This approximately 5 kb fragment is then excised from Bluescript using HindIII and BamHI and ligated into a HindIII-BamHI cut and gel purified back-bone of the pSV2CAT vector, for transfection into suitable cell lines.

5        Suitable cell lines include cell lines that have been shown to contain high levels of OP-1 mRNA, indicating that the OP-1 promoter is active in the cells. Two of these cell lines are mouse inner medullary collecting duct (IMCD) cells, and the rat bladder carcinoma line (NBT II). However other cell lines of the  
10      10 uro-genital system that produce high levels of the OP-1 message can be used in addition to the many previously mentioned cell types and cell lines.

15      The transfection of this vector into an OP-1 producing cell line is accomplished following standard techniques, i.e., transfection using calcium phosphate, liposome mediated transfection, electroporation, or DEAE-dextran transfection.

20      The transfected cells are harvested 48-72 hours after transfection with the CAT expression vector and extracts are made by successive freeze-thawing. 2  $\mu$ l of 200  $\mu$ Ci/ml  $^{14}$ C-chloramphenicol (35 to 55 mCi/mmol), 20  $\mu$ l of 4 mM acetyl CoA, 32.5  $\mu$ l of 1 M Tris-HCl, pH 7.5, and 75.5  $\mu$ l of water is added to 20 ml of cell extract, and incubated for 1 hour at 37 degrees Celsius. Upon completion of incubation, 1 ml ethyl acetate is added to the reaction, microcentrifuged for 1 minute and the top layer is  
25      removed. This top layer is dried down in a SpeedVac for 45 minutes, and each sample is resuspended in 30 ml of ethyl acetate. The samples are spotted onto a plastic-backed TLC sheet for chromatography. The thin layer is then developed in a tank containing 200 ml of 19:1 chloroform/methanol. The chromatography  
30      is run for 2 hours and placed under film for autoradiography. The activity of the  $^{14}$ C in the monoacetylated chloramphenicol series is calculated as described in Current Protocols in Molecular Biology, 1993 (Ausubel et al., eds. John Wiley & Sons, New York).

35      Upon determination of CAT activity, the main construct can be deleted in sections to determine the regions that are responsible for the observed CAT activity. Alternatively, the upstream sequences can be deleted unidirectionally, using an exonuclease

such as Bal31, and the deletion product can be analyzed in the CAT activity assay. This system can also be used in the method of the invention to screen compounds for their ability to modulate OP-1 expression by dividing the cells into several groups, and

5 culturing one group in the absence of any added compounds, and culturing the other groups with one or more candidate compound, and comparing the resulting levels of CAT activity.

While a readily assayable, well characterized, non OP-1 reporter gene is preferred in the method disclosed herein, as will

10 be appreciated by those having ordinary skill in the art, OP-1 coding sequence also may be used in the screening method of the invention. The OP-1 expression preferably is determined by an immunoassay or by Northern or dot blot or other means for measuring mRNA transcript. See, for example, WO 95/11983,

15 published May 4, 1995 for a detailed description on assaying changes in OP-1 levels in a cell or fluid.

XI. Exemplary Screening Assay for Compounds which Alter OP-1 Gene Expression in  
Endogenous Cell Type Models.

20 OP-1 is expressed in a variety of different cell types, including renal, bone, lung, heart, uterine, cardiac and neural tissue. Candidate compounds can be identified which have a modulating effect on cells of one tissue type but not another, and/or wherein the effect is modulated in the different cells.

25 The assay described below can be used to evaluate the effect of a candidate compound(s) in a particular cell type known to express OP-1 under physiological conditions.

Cell cultures of kidney, adrenals, urinary bladder, brain, or other organs, may be prepared as described widely in the

30 literature. For example, kidneys may be explanted from neonatal or new born or young or adult rodents (mouse or rat) and used in organ culture as whole or sliced (1-4 mm) tissues. Primary tissue cultures and established cell lines, also derived from kidney, adrenals, urinary, bladder, brain, mammary, or other tissues may

35 be established in multiwell plates (6 well or 24 well) according to conventional cell culture techniques, and are cultured in the absence or presence of serum for a period of time (1-7 days). Cells may be cultured, for example, in Dulbecco's Modified Eagle medium (Gibco, Long Island, NY) containing serum (e.g., fetal calf

serum at 1%-10%, Gibco) or in serum-deprived medium, as desired, or in defined medium (e.g., containing insulin, transferrin, glucose, albumin, or other growth factors).

Samples for testing the level of OP-1 production includes

5 culture supernatants or cell lysates, collected periodically and evaluated for OP-1 production by immunoblot analysis (Sambrook et al., eds., 1989, Molecular Cloning, Cold Spring Harbor Press, Cold Spring Harbor, NY), or a portion of the cell culture itself, collected periodically and used to prepare polyA+ RNA for RNA

10 analysis. To monitor de novo OP-1 synthesis, some cultures are labeled according to conventional procedures with an <sup>35</sup>S-methionine/<sup>35</sup>S-cysteine mixture for 6-24 hours and then evaluated to OP-1 synthesis by conventional immunoprecipitation methods.

15 XII. Exemplary *In vivo* Animal Model for Testing Efficacy of Compounds to Modulate OP-1 Expression

It previously has been demonstrated that OP1 can effect osteoporosis on the standard ovariectomized rat model, as indicated by the dose-response increase in alkaline phosphate and

20 osteocalcin levels following injection with OP-1. The osteoporotic rat model provides an *in vivo* model for evaluating the efficacy of a candidate modulating compound. In order to determine the effect of a candidate morphogen stimulating agent on OP-1 production and, thereby, on bone production *in vivo*, alkaline

25 phosphate and osteocalcin levels are measured under conditions which promote osteoporosis, e.g., wherein osteoporosis is induced by ovary removal in rats and in the presence and absence of a candidate modulating compound. A compound competent to enhance or induce endogenous OP-1 expression should result in increased

30 osteocalcin and alkaline phosphate levels.

35 Forty Long-Evans rats (Charles River Laboratories, Wilmington) weighing about 200g each are ovariectomized (OVX) using standard surgical procedures, and ten rats are sham operated. The ovariectomy of the rats produces an osteoporotic condition within the rats as a result of decreased estrogen production. Food and water are provided ad libitum. Eight days after ovariectomy, the rats, prepared as described above, are divided into three groups: (A) sham-operated rats; (B) ovariectomized rats receiving 1 ml of phosphate-buffered saline (PBS) i.v. in the

tail vein; and (C) ovariectomized rats receiving various dose ranges of the candidate stimulating agent either by intravenous injection through the tail vein or direct administration to kidney tissue.

5        The effect of the candidate compound on *in vivo* bone formation can be determined by preparing sections of bone tissue from the ovariectomized rats. Each rat is injected with 5 mg of tetracycline, which will stain the new bone (visualized as a yellow color by fluorescence), on the 15th and 21st day of the 10 study, and on day 22 the rats are sacrificed. The body weights, uterine weights, serum alkaline phosphate levels, serum calcium levels and serum osteocalcin levels then were determined for each rat. Bone sections are prepared and the distance separating each tetracycline staining is measured to determine the amount of new 15 bone growth. The levels of OP-1 in serum following injection of the candidate agent also can be monitored on a periodic basis using, for example, the immunoassay described in sections V and VII above.

V.       Exemplary Determination of OP-1 Protein Production 20        Where OP-1 acts as the reporter gene, detection of the gene product readily can be assayed using antibodies specific to the protein and standard immunoassay testings. For example, OP-1 may be detected using a polyclonal antibody specific for OP-1 in an ELISA, as follows.

25       1 $\mu$ g/100  $\mu$ l of affinity-purified polyclonal rabbit IgG specific for OP-1 is added to each well of a 96-well plate and incubated at 37°C for an hour. The wells are washed four times with 0.167M sodium borate buffer with 0.15 M NaCl (BSB), pH 8.2, containing 0.1% Tween 20. To minimize non-specific binding, the 30 wells are blocked by filling completely with 1% bovine serum albumin (BSA) in BSB and incubating for 1 hour at 37°C. The wells are then washed four times with BSB containing 0.1% Tween 20. A 100  $\mu$ l aliquot of an appropriate dilution of each of the test samples of cell culture supernatant is added to each well in triplicate and incubated at 37°C for 30 min. After incubation, 35 100  $\mu$ l biotinylated rabbit anti-OP-1 serum (stock solution is about 1 mg/ml and diluted 1:400 in BSB containing 1% BSA before use) is added to each well and incubated at 37°C for 30 min. The wells are then washed four times with BSB containing 0.1% Tween

20. 100  $\mu$ l streptavidin-alkaline (Southern Biotechnology  
Associates, Inc. Birmingham, Alabama, diluted 1:2000 in BSB  
containing 0.1% Tween 20 before use) is added to each well and  
incubated at 37°C for 30 min. The plates are washed four times  
5 with 0.5M Tris buffered Saline (TBS), pH 7.2. 50 $\mu$ l substrate  
(ELISA Amplification System Kit, Life Technologies, Inc.,  
Bethesda, MD) is added to each well incubated at room temperature  
for 15 min. Then, 50  $\mu$ l amplifier (from the same amplification  
system kit) is added and incubated for another 15 min at room  
10 temperature. The reaction is stopped by the addition of 50  $\mu$ l  
0.3 M sulphuric acid. The OD at 490 nm of the solution in each  
well is recorded. To quantitate OP-1 in culture media, a OP-1  
standard curve is performed in parallel with the test samples.

VI. Exemplary Production of OP-1 Polyclonal and Monoclonal  
15 Antibody

Polyclonal antibody for OP-1 protein may be prepared as  
follows. Each rabbit is given a primary immunization of 100  
 $\mu$ g/500  $\mu$ l E. coli produced OP-1 monomer (amino acids 328-431 in  
SEQ ID NO:5) in 0.1% SDS mixed with 500  $\mu$ l Complete Freund's  
20 Adjuvant. The antigen is injected subcutaneously at multiple  
sites on the back and flanks of the animal. The rabbit is boosted  
after a month in the same manner using incomplete Freund's  
Adjuvant. Test bleeds are taken from the ear vein seven days  
later. Two additional boosts and test bleeds are performed at  
25 monthly intervals until antibody against OP-1 is detected in the  
serum using an ELISA assay. Then, the rabbit is boosted monthly  
with 100  $\mu$ g of antigen and bled (15 ml per bleed) at days seven  
and ten after boosting.

Monoclonal antibody specific for OP-1 protein may be  
30 prepared as follows. A mouse is given two injections of E. coli  
produced OP-1 monomer. The first injection contains 100 $\mu$ g of OP-1  
in complete Freund's adjuvant and is given subcutaneously. The  
second injection contains 50  $\mu$ g of OP-1 in incomplete adjuvant and  
is given intraperitoneally. The mouse then receives a total of  
35 230  $\mu$ g of OP-1 (amino acids 307-431 in SEQ ID NO:5) in four  
intraperitoneal injections at various times over an eight month  
period. One week prior to fusion, both mice are boosted  
intraperitoneally with 100  $\mu$ g of OP-1 (307-431) and 30  $\mu$ g of the  
N-terminal peptide (Ser<sub>293</sub>-Asn<sub>309</sub>-Cys) conjugated through the added

cysteine to bovine serum albumin with SMCC crosslinking agent. This boost was repeated five days (IP), four days (IP), three days (IP) and one day (IV) prior to fusion. The mouse spleen cells are then fused to myeloma (e.g., 653) cells at a ratio of 1:1 using 5 PEG 1500 (Boeringer Mannheim), and the cell fusion is plated and screened for OP-1-specific antibodies using OP-1 (307-431) as antigen. The cell fusion and monoclonal screening then are according to standard procedures well described in standard texts widely available in the art.

10 VII. Exemplary Process for Detecting OP-1 in Serum

Presented below is a sample protocol for identifying OP-1 in serum. Following this general methodology OP-1 may be detected in body fluids, including serum, and can be used in a protocol for evaluating the efficacy of an OP-1 modulating compound *in vivo*.

15 A monoclonal antibody raised against mammalian, recombinantly produced OP-1 using standard immunology techniques well described in the art and described generally in example VI., above, was immobilized by passing the antibody over an agarose-activated gel (e.g., Affi-Gel™, from Bio-Rad Laboratories, 20 Richmond, CA, prepared following manufacturer's instructions) and used to purify OP-1 from serum. Human serum then was passed over the column and eluted with 3M K-thiocyanate. K-thiocyanante fractions then were dialyzed in 6M urea, 20mM PO<sub>4</sub>, pH 7.0, applied to a C8 HPLC column, and eluted with a 20 minute, 25-50% 25 acetonitrile/0.1% TFA gradient. Mature, recombinantly produced OP-1 homodimers elute between 20-22 minutes, and are used as a positive control. Fractions then were collected and tested for the presence of OP-1 by standard immunoblot using an OP-1 specific antibody. Using this method OP-1 readily was detected in human 30 serum. See also, PCT/US92/07432 for a detailed description of the assay.

IX. Considerations for Formulations and Methods for Administering Therapeutic Agents

Where the OP-1-modulating agent identified herein comprises 35 part of a tissue or organ preservation solution, any commercially available preservation solution may be used to advantage. For example, useful solutions known in the art include Collins

solution, Wisconsin solution, Belzer solution, Eurocollins solution and lactated Ringer's solution. Generally, an organ preservation solution usually possesses one or more of the following properties: (a) an osmotic pressure substantially equal to that of the inside of a mammalian cell, (solutions typically are hyperosmolar and have K<sup>+</sup> and/or Mg<sup>++</sup> ions present in an amount sufficient to produce an osmotic pressure slightly higher than the inside of a mammalian cell); (b) the solution typically is capable of maintaining substantially normal ATP levels in the cells; and (c) the solution usually allows optimum maintenance of glucose metabolism in the cells. Organ preservation solutions also may contain anticoagulants, energy sources such as glucose, fructose and other sugars, metabolites, heavy metal chelators, glycerol and other materials of high viscosity to enhance survival at low temperatures, free oxygen radical inhibiting agents and a pH indicator. A detailed description of preservation solutions and useful components may be found, for example, in US Patent No. 5,002,965.

Where the OP-1-modulating agent is to be provided to an individual, e.g., the donor prior to harvest, or the recipient prior to or concomitant with transplantation, the therapeutic agent may be provided by any suitable means, preferably directly (e.g., locally, as by injection to the tissue or organ locus) or systemically (e.g., parenterally or orally).

Useful solutions for parenteral administration may be prepared by any of the methods well known in the pharmaceutical art, described, for example, in Remington's Pharmaceutical Sciences (Gennaro, A., ed.), Mack Pub., 1990. Formulations may include, for example, polyalkylene glycols such as polyethylene glycol, oils of vegetable origin, hydrogenated naphthalenes, and the like. Formulations for direct administration, in particular, may include glycerol and other compositions of high viscosity to help maintain the agent at the desired locus. Biocompatible, preferably bioresorbable, polymers, including, for example, hyaluronic acid, collagen, tricalcium phosphate, polybutyrate, lactide and glycolide polymers and lactide/glycolide copolymers, may be useful excipients to control the release of the agent in vivo.

As will be appreciated by those skilled in the art, the concentration of the compounds described in a therapeutic composition will vary depending upon a number of factors, including the dosage of the drug to be administered, the chemical 5 characteristics (e.g., hydrophobicity) of the compounds employed, and the route of administration. Where the morphogen-stimulating agent is part of a preservation solution, the dosage likely will depend for example, on the size of the tissue or organ to be transplanted, the overall health status of the organ or tissue 10 itself, the length of time between harvest and transplantation (e.g., the duration in storage), the frequency with which the preservation solution is changed, and the type of storage anticipated, e.g., low temperature. In general terms, preferred ranges include a concentration range between about 0.1 ng to 100 15  $\mu$ g/kg per tissue or organ weight per day.

Where the therapeutic agent is to be administered to a donor or recipient, the preferred dosage of drug to be administered also is likely to depend on such variables as the type and extent of progression of the disease, the overall health status of the 20 particular patient, the relative biological efficacy of the compound selected, the formulation of the compound excipients, and its route of administration. In general terms, a suitable compound of this invention may be provided in an aqueous physiological buffer solution containing about 0.001% to 10% w/v 25 compound for parenteral administration. Typical dose ranges are from about 10 ng/kg to about 1 g/kg of body weight per day; and preferred dose range is from about 0.1  $\mu$ g/kg to 100 mg/kg of body weight per day.

The invention may be embodied in other specific forms without departing from the spirit or essential characteristics thereof. 30 The present embodiments are therefore to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims rather than by the foregoing description, and all changes which come within the 35 meaning and range of equivalency of the claims are therefore intended to be embraced therein.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: OZKAYNAK, ENGIN  
OPPERMANN, HERMANN

(ii) TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING  
MORPHOGENIC PROTEIN EXPRESSION

(iii) NUMBER OF SEQUENCES: 7

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: PATENT ADMINISTRATOR, CREATIVE BIOMOLECULES  
INC.  
(B) STREET: 45 SOUTH STREET  
(C) CITY: HOPKINTON  
(D) STATE: MA  
(E) COUNTRY: USA  
(F) ZIP: 07148

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
(B) COMPUTER: IBM PC compatible  
(C) OPERATING SYSTEM: PC-DOS/MS-DOS  
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US 07/938,021  
(B) FILING DATE: 28-AUG-1992  
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: KELLEY, ROBIN D  
(B) REGISTRATION NUMBER: 34,637  
(C) REFERENCE/DOCKET NUMBER: CRP-091

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (508)-435-9001  
(B) TELEFAX: (508)-435-0992

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17415 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: homo sapiens

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

|  |      |
|--|------|
| TCAACCGGTC TCTTTAGGTT TTGGCTGTGC TTATTACTAT TCATTCAACA GGTACTAATT  | 60   |
| GAGCACCTGC TGTGTGCCAG GCTCAGAATA GGCTCAGGTG AGATGCACAA AGAAGGGTAA  | 120  |
| ACTAGAAATCC TTGCTTAGAC ACTGACGGAT CAGTTGTTTC ATATGTAAAT TGTAGCACCA | 180  |
| AGACCTGCTG CCCCTGCCCC CAGCCTCACC TGCTTGTGAA GATCCCTCCA AAAGATTGAA  | 240  |
| GAGTAGATAA AAAGCAGAGA CTACTACTGA AGAACAGGGC TGCTTGGCT CCTTATTATT   | 300  |
| TCAGACTTTG GAAGAAAATG ACCTCCTTT TCTCTACTGG CACTGAGTGC ATAGCTGACC   | 360  |
| TAGCAAGCCA GGCCCTGGAGG GCGTGTGCAG GGCTGGGAC CGAGCCTGGT TTCTGTTCCC  | 420  |
| TGCTCTGCAG CTCAAGCACT TGCTGTTCCCT CCACCTGGGA TGCCTTCCCC TGGAAAAGCC | 480  |
| TGTCTCTTTC TTGCTTTCA GGACTCAGGT CAGTGGCATC TCCTCCAAAA ACTCCCCCTTC  | 540  |
| CCACCCCTCCA TCACCTCACC CTGTTTATCT GCGCCCCCGC CCCCAC TGTCACTTAT     | 600  |
| TGCAGGCTGA AGTGAACCCAG GCTCTCCAGT TGTACACTCT CAGATGGACC CTGGACGACT | 660  |
| GTGGCACTCC TGCAATTTC CCAGTCTCCC TGGGGTAGGA TTCCTGCTTG CCAGGATGCC   | 720  |
| CACCTTCCT TCTCCCTCCT GCATGTCTC CTCTGCCTGG CTTCTGAATT GTTTCCAGAG    | 780  |
| AGAGTGATAG ACAAGATCTG CCTCTCCCTC AGTCCCTGAA TCTTATTTAA GGCTCTTGCT  | 840  |
| TTGCTTCCTT GGCCCTGGAGG CGGCTCCTTG ATGGAGTCTG CCATGTGGGT TCGCTCATGG | 900  |
| CCATGTCTTC CTGCCAGCA TGGTGCTTGG CCCTGGGACT GGCCACATAA TATCTGGGCC   | 960  |
| AGGTGCAAAA TTAGTACGGG GCAGGGGTA CTTTGTTCAT AGGTGATTCA GAACCACATA   | 1020 |
| TGGTGACCTC AGAGTAGGAA ACCAAGTGTG GGGCCCTTAA GAGCTGGGG GCCCTGTACG   | 1080 |
| ACTGTCCAGG TTGCAGGCC CACAGCTCGC CTCTGATAT CCTGTGCTCC ATGCTTGTCT    | 1140 |
| GTTGAAGGAA GGAGTGAATG GATGAAGAGC AGGTGGTGGG GGGTGGTTTG AGGGCCTTGC  | 1200 |
| TGGTGGGTGG GTAGAGGCC CTCCCTGGCA TGGGGCTCAA GACCTGTTCC ATCCCACAGC   | 1260 |
| CTGGGGCTGT GTGTAAATGG CCAGGACCTG CAGGCTGGCA TTTTCTGCT CCTTGCTGG    | 1320 |
| CTCTGGCTCC CCTTCTCCA CCCATGTGGC CCCTCAGGCT GCCATCTAGT CCAAAAGTCC   | 1380 |
| CAAGGGAGAC CCAGAGGCCA CTTGGCAAAC TACTTCTGCT CCAGAAAAT GTAGAAGACC   | 1440 |

|             |            |            |             |             |            |      |
|-------------|------------|------------|-------------|-------------|------------|------|
| ATAATTCTCT  | TCCCCAGCTC | TCCTGCTCCA | GGAAGGACAG  | CCCCAAAGTG  | AGGCTTAGCA | 1500 |
| GAGCCCCCTCC | CAGACAAGCG | CCCCCGCTTC | CCCAACCTCA  | GCCCTTCCCA  | GTTCATCCCC | 1560 |
| AAGGCCCTCT  | GGGGACCCAC | TCTCTCACCC | AGCCCCAGGA  | GGGAAGGAGA  | CAGGATGAAC | 1620 |
| TTTTACCCCG  | CTGCCCTCAC | TGCCACTCTG | GGTGCAGTAA  | TTCCCTTGAG  | ATCCCACACC | 1680 |
| GGCAGAGGGA  | CCGGTGGGTT | CTGAGTGGTC | TGGGGACTCC  | CTGTGACAGC  | GTGCATGGCT | 1740 |
| CGGTATTGAT  | TGAGGGATGA | ATGGATGAGG | AGAGACAGGA  | GAGGAGGCCG  | ATGGGGAGGT | 1800 |
| CTCAGGCACA  | GACCCTTGGA | GGGGAAGAGG | ATGTGAAGAC  | CAGCGGCTGG  | CTCCCCAGGC | 1860 |
| ACTGCCACGA  | GGAGGGCTGA | TGGGAAGCCC | TAGTGGTGGG  | GCTGGGTGT   | CTGGTCTCAG | 1920 |
| GCTGAGGGGT  | GGCTGGAAAG | ATACAGGGCC | CCGAAGAGGA  | GGAGGTGGGA  | AGAACCCCCC | 1980 |
| CAGCTCACAC  | GCAGTTCACT | TATTCACTCA | ACAAATCGTG  | ACTGCGCACG  | TACAGTGGCT | 2040 |
| ACCAGGCCT   | GGGTTCAAGG | CACTGCGGTT | ACCAGAGGTG  | CGGAGAAAGAT | CGCTGATCCG | 2100 |
| GGCCCCAGTG  | CTCTGGGTGT | CTAGCGGGGG | TAAGAAGGCA  | ATAAAGAAGG  | CACGGAGTAA | 2160 |
| CTCAAACAGC  | AATTCCAGAC | AGCAAGAGAA | ACTACAGGAA  | AGAAAACAAA  | CGTGCAGGG  | 2220 |
| GCGAGGCGAG  | GAAACAACCT | CAGCTTGGCA | GGTCTTGGAG  | GTCTCTGGGA  | GGAGAAAGCA | 2280 |
| CGCTCTGATG  | GGGGCGGGAG | GTGGTGAGTG | GGGAGAGGTC  | CAGGCAGGAGG | GAATGGCGAG | 2340 |
| CGAGAGACAG  | GCTGGCAACG | GCTTCAGGGA | GGCGCGGAGG  | GGTCAGCGTG  | GCTGGCTTAA | 2400 |
| AAGGATACAT  | GGGACTAGGG | GCAAGACCGG | CTCAAGGTCA  | CCGCTTCCAG  | GACCTTCTAT | 2460 |
| TTCCCGGCCA  | CCTCCCGCCT | CCCCCAACTT | TTCCCACCGC  | GGTCCCGCAGC | CCACCCGTCC | 2520 |
| TGCTCGGGCC  | GCCTTCCTGG | TCCGGACCGC | GAGTGCCGAG  | AGGGCAGGGC  | CGGCTCCGAT | 2580 |
| TCCTCCAGCC  | GCATCCCCGC | GACGTCCCCG | CAGGCTCTAG  | GCACCCCGTG  | GGCACTCAGT | 2640 |
| AAACATTGT   | CGAGCGCTCT | AGAGGGAATG | AATGAACCCA  | CTGGGCACAG  | CTGGGGGGAG | 2700 |
| GGCGGGGCCG  | AGGGCAGGTG | GGAGGCCGCC | GGCGCGGGAG  | GGGCCCCCTCG | AAGCCCGTCC | 2760 |
| TCCTCCTCCT  | CCTCCTCCGC | CCAGGCCCCA | GCGCGTACCA  | CTCTGGCGCT  | CCCGAGGCAG | 2820 |
| CCTCTTGTGC  | GATCCAGGGC | GCACAAGGCT | GGGAGAGCGC  | CCCCGGGGCC  | CTGCTATCCG | 2880 |
| CGCCGGAGGT  | TGGAAGAGGG | TGGGTTGCCG | CCGCCCCGAGG | GCGAGAGCGC  | CAGAGGAGCG | 2940 |
| GGAAGAAGGA  | GCGCTCGCCC | CCCCGGCTGC | CTCCTCGCTG  | CCTCCCCGGC  | GTTGGCTCTC | 3000 |
| TGGACTCCTA  | GGCTTGCTGG | CTGCTCCTCC | CACCCGCGCC  | CGCCTCCTCA  | CTCGCCTTTT | 3060 |

|             |             |             |             |             |             |      |
|-------------|-------------|-------------|-------------|-------------|-------------|------|
| CGTTCGCCGG  | GGCTGCTTTC  | CAAGCCCTGC  | GGTGCGCCCG  | GGCGAGTGCG  | GGCGAGGGG   | 3120 |
| CCCGGGGCCA  | GCACCGAGCA  | GGGGCGGGG   | GTCCGGGCAG  | AGCGCGGCCG  | GCCGGGGAGG  | 3180 |
| GGCCATGTCT  | GGCGCGGGCG  | CAGCGGGGCC  | CGTCTGCAGC  | AAGTGACCGA  | GCGGCGCGAC  | 3240 |
| GGCCGCCTGC  | CCCCCTCTGCC | ACCTGGGGCG  | GTGCGGGCCC  | GGAGCCCGGA  | GCCCCGGTAG  | 3300 |
| CGCGTAGAGC  | CGGCAGCGATG | CACGTGCAGCT | CACTGCGAGC  | TGCGCGCCG   | CACAGCTTCG  | 3360 |
| TGGCGCTCTG  | GGCACCCCTG  | TTCCCTGCTGC | GCTCCGCCCT  | GGCCGACTTC  | AGCCTGGACA  | 3420 |
| ACGAGGTGCA  | CTCGAGCTTC  | ATCCACCGGC  | GCCTCCGCAG  | CCAGGAGCGG  | CGGGAGATGC  | 3480 |
| AGCGCGAGAT  | CCTCTCCATT  | TTGGGCTTGC  | CCCACCGCCC  | GCGCCCGCAC  | CTCCAGGGCA  | 3540 |
| AGCACAACTC  | GGCACCCATG  | TTCATGCTGG  | ACCTGTACAA  | CGCCATGGCG  | GTGGAGGAGG  | 3600 |
| CGGGCGGGCC  | CGGCAGGCCAG | GGCTTCTCCT  | ACCCCTACAA  | GGCCGTCTTC  | AGTACCCAGG  | 3660 |
| GCCCCCCTCT  | GGCCAGCCTG  | CAAGATAGCC  | ATTTCCTCAC  | CGACGCCGAC  | ATGGTCATGA  | 3720 |
| GCTTCGTCAA  | CCTCGGTGAG  | TAAGGGCAGG  | CGAGGGTACG  | CCGTCTCCTT  | TCGGGGGCAC  | 3780 |
| TTTGAGACTG  | GGAGGGAGGG  | AGCCGCTTCT  | TCTATGCAGC  | CCGCCCAGCT  | TTCCGCTCCT  | 3840 |
| GGCTGAAATC  | GCAGTGCTG   | CCCGAGGGTC  | TCCCACCCAC  | AGCCCTATGA  | CTCCCAAGCT  | 3900 |
| GTGTGCGCCC  | CCAGGTCGGG  | CCGCTGGTC   | GGTGAGCCTG  | TAGGGGTTAC  | TGGGAAGGAG  | 3960 |
| GGATCCTCCG  | AAGTCCCCTC  | CATGTTACGC  | CGCCGGCCGC  | ATCTCTGGGG  | CTGGAGGCAA  | 4020 |
| GGGC GTTCAA | AGCGGGGGC   | TCGGTCATGT  | GAGCTGTCCC  | GGGCCGGCGC  | CGGTCCGTGA  | 4080 |
| CCTGGATGTA  | AAGGGCCCTT  | CCCAGGAGG   | CTGCCTTGCC  | GCCCTTCCTG  | GGCCCCCTCTC | 4140 |
| AGCCCTGCCT  | GGCTCTGGCA  | TCGCGGCCGT  | CGCACCCCCCT | TACCCCTCCCT | GTCAAGCCCT  | 4200 |
| ACCTGTCCCC  | TCGTGGTGCG  | CCCGCCTTAG  | GCTACCGGCC  | GCTCCGAGCC  | TTGGGGCCCC  | 4260 |
| TCTCCGGCG   | CCGATGCCCC  | ATTCTCTCTT  | GGCTGGAGCT  | GGGAAGAAA   | CGGTGCCATT  | 4320 |
| GCTAATTTTC  | TTTGTTCCT   | TTCTTTGTTT  | ATTTTTTCT   | TTTTTCTTTT  | TTTTTCTTTT  | 4380 |
| CTTTTCTTTT  | CTTTTTTTTT  | TTTTTGAGA   | CGGAGTTCA   | CTCTTGCTCG  | CCCAGACTGG  | 4440 |
| AGTGCATGG   | CGCGATCTCT  | GCTCACCGCA  | ACCTCTGCCT  | CCCGGGTTCA  | AGCGATTCTC  | 4500 |
| GTGCCTCAGC  | CTCCCGAGTA  | GCTGGGATTA  | CAGGCATGCG  | CACCATGCCT  | GGCTAATTTT  | 4560 |
| GTATTTAGT   | AGAGACAGGG  | TTTCTCCATG  | TTAGGCAGGC  | TGGTCTCGAA  | CTCCCGATCT  | 4620 |
| CAGGTGATCC  | TCCCGCCTCA  | GCCTCCAAA   | GTGGTGCTGG  | GATTACAGGC  | GTGAAGCTGT  | 4680 |

|   |      |
|---|------|
| GCCCTGCCGC TAGTCTTCTA TTTTAAGTAT TTAGTGGTAG GTCCCGGGCC GGCAGAACATCT | 4740 |
| ATTTTCAGCA TTTACCACGT GTGGCGCGCA AACCAACAGGT TTTGGCGATT GGGTTGCGCG  | 4800 |
| GGATCTCAGA GCTGACGACC GCGGGGGCCT GGGGGTCCCG GTTTCCGACT GGAGCCGCGA   | 4860 |
| CGACCCCCGGC GACGGCAGCC TGGGGCTGCA GCCGAGGGCC GGGGAGCTCC CCCTCCATAT  | 4920 |
| GTGCGCGCAC ATTCTCCAGA CTTGCTCAAA CTAACCCCCC GGAGCAGCGC ACGGGCTGGG   | 4980 |
| ACTGATGATC AAATATTGAG TTTCCGAGAT AACACACCCC GATAGCGCTG TTTCCTGAGC   | 5040 |
| CGCTTTCATT CTACTTGTGT AACTTGCTGC GAAAACCGA ACCAAGTCAA GACAGCAAAC    | 5100 |
| TCACGCCAC GGGCCTGTGT CAACATGGAA ATAATGATAC TGAAGCCCCA CGCTGGGCAC    | 5160 |
| CTGGGGCGTG GACTGGGGC GCGGGGAAG CGCAGATCCG CCTTCATGCT TCCCCCTCCTC    | 5220 |
| CTGATAAGGT CCCTGGAGTT CCCGGGAGCC ATTGTCTGTA CTTAATAATA ACTAAATCCA   | 5280 |
| ACTAGTGAAC CAAGCTTCAG CGAGGCAAGG GGAGGGAGGT TTAGATGCCA AAATTACCTT   | 5340 |
| CAAAAAAGTT TAAATTATAC TAAGCAGCCA GTTAAGAAGG AAGCAGCAAT ATATGACCTG   | 5400 |
| ATTTAGAACC ATCTCCAAGA TGTATGAGGT GGAAAGAAGC AAGGTGCAGA TGAGTGGGCT   | 5460 |
| GCATGTGTGC TTGTATATCA TCGTGTCTC CTGGAGGAAG ACACCAGGAA CTGGAGAGAG    | 5520 |
| ATTTTACTGG AGGGGTATAT CGCGGGGGCA TAGCTGGGGC TTACGGACTG GGACGTGGGG   | 5580 |
| TCTGATTTTT CGTCGTCTGC ACTTCTGTAT TTGTGATTTT TTTAAAACAA TGTGTATTTA   | 5640 |
| TTAACTATAC CAAAAAATAA AGGAAAATTC CAAATACATA CATATAAATA ATGAACCGCA   | 5700 |
| GAGCTCTGTC GCCCTCCTGA AGCCTGGGT TAGCCAGGGC CCTTTCTCTG GTGGGGGATT    | 5760 |
| TATAGCATCT TCCCTTCTGT TGGGTACCCC GGACTCCCAC TGAATGTGCA GGTCCCAGTG   | 5820 |
| GCTGCCTTCA GAGCCTGGCT CGAACATCTTA AAAAGGTATT TGTAACTCTCT GGCTTCTGCA | 5880 |
| GAAGGCCCTG CAAACCAAGA GCAAAAAGC CCCCAGTGCT TATGGGCCGG CAGTGTGGGC    | 5940 |
| TAGGCCCGGG GCTCCCTGTC CCCAAGAGAA AGACCAGGTT GCTCGGAGGG TGCCCTCTGGG  | 6000 |
| AACTTTGGTG CGGGCTATTT GCTCCCCCA TGGCGGCAGG AGCAAGCTGG GACTTGTGG     | 6060 |
| GGAAGGCCAC AGCTGGGTGG TTTTCCTCCT CTGGCTGTAC ATACACCTTT CAATCCATT    | 6120 |
| CTTTCATCTT GAAAGGACAA AGACCGGCTT GTCTGAGCCT CTTAACAGT CAGGCTGGCT    | 6180 |
| TTGGGCTTTG GGGACCCCTGA CTTTCTCAGG TCTAGCTTTC TGGGACATCA CTCCAAATTA  | 6240 |
| GATGGCAGAG TGGCTTTAA CAGAGCGCAC TGACCTTGTT TTCTTCTCT CTCTGTCCCT     | 6300 |

|             |            |            |             |             |            |      |
|-------------|------------|------------|-------------|-------------|------------|------|
| AAACTCGAGG  | TCATTAGTTA | GGTGAAGACC | TGGGCTGCAG  | TTTGGCGAGA  | CACTTCCTGT | 6360 |
| AGATGCTTCT  | AATGTTGGCC | TTTAATTCT  | GCTAAGCAGC  | AGCACACAAA  | TAAATGGCCT | 6420 |
| GTCCCTTCTA  | TCCTGTTGTA | GCTTGGATT  | TCTCCATAGG  | AGGGACTTGG  | GGGTGGCAGT | 6480 |
| AGGGTTGGAG  | AGGGTTGGGG | GGAGGTGTAG | GAGACTTGTC  | TGGCCACTGA  | GTGGCTGAG  | 6540 |
| AAAGTACTGC  | TATAGTGT   | TTCTTGGAT  | TGCAAATCAT  | GTTGATCTGA  | ACTGCTGATT | 6600 |
| TGAAAGTGGAT | TGAGAGGATG | GAACAATAGA | AGGAGGATAT  | GGCTCAGGAC  | AGTCAAGTAC | 6660 |
| TGGAAGAGGG  | AAAGGTACAA | AGAGGTGTTG | GCACTGAATG  | ACCCCTGAACA | GGGCTGCCCT | 6720 |
| GGAAATATCA  | GAGGTGAGTG | ACAAAGAGAA | CTCTAGTCGA  | AGGTCTGGAA  | GTCAATTATT | 6780 |
| GTCTCCAGCT  | TTTGTCCCAC | CCTAAGGGAT | GGAGCATGAA  | CTTCATGCAT  | CTAACATCCC | 6840 |
| TCCAGGAGCG  | CTGAGGTTCT | GGGAATTCCC | AGTGCTGGCT  | ACCATGCCAT  | TCTTTCTCA  | 6900 |
| TTCACTCAAG  | AGCGTATTGG | GATATGCGTG | CATGAAAGCA  | ATGTAATTAT  | GGGCACAACC | 6960 |
| TCAAAACCTG  | CTCTAATT   | TTTTTTTTT  | GGAGATGGAG  | TCTCGCTCCA  | TCACCCAGGC | 7020 |
| TGGAGTGC    | GGCGCGATC  | TCAGCTCACT | GCAAGCTCAG  | ACCTCCAGGG  | TTCACACCAT | 7080 |
| TCTCCTGCCT  | CAGCCTCCCG | AGTAGCTGGG | AATACAGGGCG | CCCGCACCAT  | GCGCGGCTAA | 7140 |
| TTTTTTGTA   | TTTTTAGTAG | AGACGGGTT  | TCACTGTGTT  | AGCCAGGATG  | GTCTCGATCT | 7200 |
| CCTGACCTCG  | TGATCCACCC | GCCTCGGCCT | CCCAAAGTTC  | TGGGATTACA  | GGCGTGACAG | 7260 |
| CCGTGCCCGG  | AATCTGCTCT | AATT       | AGATATCATT  | TGCAAAC     | GGGCACTTGA | 7320 |
| GTCACTCAGT  | AAGATATTAT | TTACAACCCC | ACCATAGATT  | CAAACCTCTG  | TCCTAGAATG | 7380 |
| TTGTCGAGTT  | AGGCATCTGG | CTTGCAGCAA | CAGCTGGCTT  | TCCTGTCTAT  | GCTGTCTCCT | 7440 |
| TCCAGGGAGG  | ATGTTTCACC | CTTCATATTG | AGGAAATGGG  | CACAGAGAAC  | CCATTCTCT  | 7500 |
| TACTCATCAT  | GTAACCTCAG | TGGGATGGTC | AGATCTATCT  | TTAACCTGGC  | CACTCTTCCA | 7560 |
| CAAGCTCACA  | CTGACTCCAG | CAAGATCTTA | AACTAGAAGG  | CAGGAGTTCA  | AATCCTAGCT | 7620 |
| GGTGCAGTGG  | CCAAATCTCG | GCTCACAGCA | CCTTCTGCCT  | CCTGGGCTCA  | AGCGATCCTC | 7680 |
| TGACCTCAGT  | CTCCCAAGTA | GCTGGGACCA | TAGGCATGCA  | CCACTATGCC  | TGGCTAATT  | 7740 |
| TTGTATT     | GTAATT     | GTAGAGACAG | AGTTTCACCA  | TGTTGCCAG   | CCCAGTCTTG | 7800 |
| AACTCCTGGA  | CTCAAGCAAT | CTTCCCACCT | TTGCCTACCA  | GAGTGCCGGG  | ATTACAGGTG | 7860 |
| TGAGCCATCA  | TGCTAGTTGC | GCACAGTTGG | GCGAAACTGA  | CAGATGAGAA  | AGCAGAACCT | 7920 |

|   |      |
|---|------|
| CGTGAGTCCA CTCAGTAAGA GACTCCCTAC TTTCTTCTG AGTCTTGTT TCTCATCAAT     | 7980 |
| TGAATGGCAA TAAACAACTT GGTGGCCCAA GAGTTGATGA CAACAGTCCT ATAAGATTAT   | 8040 |
| ACATGTAAAA GAAACAGAGT ATTCTACAAA TATCAGTTAT TGATAGTTCA ATAGGCAACC   | 8100 |
| TGACATTACC TTTCTTGGG ACTTGATGAA CAACTCAGAA ACTCATTAAT ATCAAACCCA    | 8160 |
| ATGGTGAGCA CTTGGTCTTT ATTTATGGCT GTAAGAGAAG AAATTGAATT AACTCTATGT   | 8220 |
| AAATGCCAAC TAAGAACATC GAAGTCTGAA ATCAACAGTT TTCCTCGCTC ATACGACACA   | 8280 |
| CCCAAACCTCA AGCAGTGGTT CCAAGCCCT TTGGAAAATA CCATGGGCTA ACGACTTTAA   | 8340 |
| AAGCTTAGAA GTGAATTCTA CTTACTTATT ACTTAAAAGT GGTTCTCAAA CTTCAAGGTG   | 8400 |
| AATCAAAATC ATCTGTAGAG CTTGTTAAA CACAGGTTGC TGGTCCACCC CAAGAGTGTC    | 8460 |
| TTGAGTCAGT AGGTCTCAAG TAGGGCTCAA GAATATGCAT TTCTAATGAG CTCCAGGTGA   | 8520 |
| GTCTAAGTGT TAGTCGTCGG TCTTGGGACC ACAACTTTGG GAACAATTGA TTTAGAAGAA   | 8580 |
| CTCAAAGATC AGAAAGGGT GGAATATTTT TAAAATTGTG GTAAAATACG CATAAACAGA    | 8640 |
| AAAGGTACAA TTTTAACCAC TTAGAGAGAG GTGGGATCTA AGAACAGAAA TTGTTATGCC   | 8700 |
| ATCAAAGGTG AGTTCAGATA AGCATTATTA AATGGTATCT ATGGATAAAC TTCAAGGGCC   | 8760 |
| CTGTGGAGCA ACCCAATGCT GGGATGGGT CCAGGTGTGC TATGGTTTGG ATGTGGTTTG    | 8820 |
| TCCCTACAAA AACTCATGTT GAAATTAAAT TGCCAGTGTACACATTATTGA GAGGTTATGG   | 8880 |
| ACTTTAAGA GGCATTTGGG TCATGAGGGG TCCACCTTCA GGGATTAGTG CAGTCTCCAG    | 8940 |
| GGAGTGAGTG AGTTCCCATT CTAGTGGGAC TGGATTAGTT ACCATACAGT GGTTGTTATA   | 9000 |
| AAGTGAGGCT GCTTCTGGTG TTTTATCTGT TTGCAGGCAC TTCCCTCCCC TTCCACTTCT   | 9060 |
| CTGCCAGGTT AGGATGCAGC ATGAGGCCCT CACCAGAAC TGACCAGATG TGGCTGCCTG    | 9120 |
| ATCTTGAAC TCCCCAGTCCC CAGAACCATG AGCTAAATAA ACCTTTTTC TCTATAAAATT   | 9180 |
| ACGCAGTCTA GAGTATTCTA TTATAGCAAC ACAAGACAGA CTAAGACACA GTGGTAGAAA   | 9240 |
| GAACACTACT GACTTCTCCC ATACTCTGGC CTATGGACAA GAGTGACAGA CAGACAAGAG   | 9300 |
| TGAATATCAG GGCCCTCAGG CACATTCCCTC TCTGCCCTT CCTCCCTTCT TGCAGAGTCT   | 9360 |
| CCAGTGACTG CCAGCTAACATG CTATCATAGA CCCCACCTTT CCCCTGACTT GATTGGACCA | 9420 |
| GAAGCAGCCT CCTGATCCAT GGCCAACAAT CAGATTCACT TTCAAGAATT TGAACTAAGA   | 9480 |
| GACACTAGGA AGATGGCCCT TGAGCTGTGA GTCCTACACT TGAAAGTTCT TAGCATCTTG   | 9540 |

|  |       |
|--|-------|
| GTCAGGTACC CACCAGGGCC ATGTGCAAAC TGAGATAATG GGGACATGGA ACAAGGGTAA    | 9600  |
| GTGGAGAGGG CTGGCTGGAG AGAGACGGGC AGAGGAAAGC CCTGCCAAGA GGAGCAGAGA    | 9660  |
| TGAGAGACCT TGGAGGGAGA GGTAATAAAA GGAGGCAAAG ATGATTTCC ATGCTTACAA     | 9720  |
| CTCACAGCTG AGGCCTAACT ATCTTATGT CCATAAGAGG CATCCTTGTG TCGAACCTCT     | 9780  |
| CCTCTTCTT GGGTCAATGG GGGATGGTTG CAAGGGACCA TCAGTAGGAA GGCATAGTAC     | 9840  |
| ACTAACCCAG TCTGGGGTGG GCTTTAGAC TAGTCTTCCT CCCATGCTCC TCCTCCCATT     | 9900  |
| GGAACCCCGG ACTTTCAAGA CTGCTACCTA GCACACCAGT GCACCAGATG TCACTCAAAA    | 9960  |
| CCTCTTCAGC AATGGCCCAC TCACCTCAA AAAGGCTGAA GAGCAGACTG GCTGGTTCT      | 10020 |
| TCATGGTGGA GGGCAGTCT GGGAGGTTT AAGGTTGAAG ATGAAAACCTT TCACTTTG       | 10080 |
| CTCAATGGTC TGAAAAAGAG AAGGACCAGC AAGTGAACTG AAGCCTCCTG GAAAGCATCT    | 10140 |
| TGATAACAGG GGCAGAGTTT CAAGATGAGA AGCTGTGGCA CTTACTCTGG CTTGGAAAT     | 10200 |
| GACCTCTAAG TATCTCAGTT AATTAAAGGA CTCAAACTCT AGACTCGAAG GAGAACATCT    | 10260 |
| ACAATTTCA ATAACATAGT CTACCCTCCC CTCCCTCCCC CACCTTCACC TCTTCTTCA      | 10320 |
| TCACAGGCTT ACAGGGCACC TCTTAGAGCC AGGCACGGTG TTGGGATCAG GAACAAGGCC    | 10380 |
| ACTGCTCACA TCCAGAGCCT GTGCTACTTA AGAAGCTTCC AGGACCTCTT GGATGGCTGT    | 10440 |
| GGTTAGTGCC CTACTTTCC CAGCAGGTTG GATGCAGAAT CATGCTCTTG TCGTTCAGGA     | 10500 |
| TGACCATGGG GACCAGGGT CTGAGCCTGT GACCCTCCAG TCTACAGTGT GTTGGTGAGG     | 10560 |
| AAGGAGCAGT TGTCACTGGG GTCACTGGCA ATGGGCATGC CTCCATCTAG CTTAGGCAAG    | 10620 |
| ATGCTTAGAC TCAGAGCCAG AGAGTGAAAC CCAGACACTA ATGAGCTGTC GGTGTTGGTG    | 10680 |
| TGTGTTCTCT TCCTCTTCCA GTGGAACATG ACAAGGAATT CTTCCACCCA CGCTACCACC    | 10740 |
| ATCGAGAGTT CCGGTTTGAT CTTTCCAAGA TCCCAGAAGG GGAAGCTGTC ACGGCAGCCG    | 10800 |
| AATTCCGGAT CTACAAGGAC TACATCCGGG AACGCTTCGA CAATGAGACG TTCCGGATCA    | 10860 |
| GCGTTTATCA GGTGCTCCAG GAGCACTTGG GCAGGTGGT GCTATACGGG TATCTGGGAG     | 10920 |
| AGGTGCTGAG TTTCCTCTGG GGGCAGAGGA AGAAGGTGGT GAGGGTTCC CTCCCCCTCCC    | 10980 |
| ACCCCATGAG CTCTGCTTCC CATCTGTTGG GGTAGTGGAG CTGTGACCTG CTAACGCGAA    | 11040 |
| GCCC GTGTCT CTCCTCCTCT CTCGCAGGGAA ATCGGATCTC TTCCCTGCTCG ACAGCCGTAC | 11100 |
| CTCTGGGCCT CGGAGGAGGG CTGGCTGGTG TTTGACATCA CAGCCACCAAG CAACCACTGG   | 11160 |

|  |       |
|--|-------|
| GTGGTCAATC CGCGGCACAA CCTGGGCCTG CAGCTCTCGG TGGAGACGCT GGATGGTGAG  | 11220 |
| TCCCCCGCCA CTGCCAGTCC TAATGCAGCC TGTGCTCCTG GACTTCAGGA GGGTCTCAGC  | 11280 |
| AGTGCTCATG CTTGCTTCAC TACAAACAGG CTTCCCGCC CCTCCCAACC AGTACTCCAT   | 11340 |
| GTTCAGCCTT TTGATCCTGC AGCCCTGTCC CGCTCGTGGC CCTCCTGTAA CTGCTCTTCT  | 11400 |
| GTGCACTTGG CTGCTTCCTG TCCAGGGCAG ACGATCAACC CCAAGTTGGC GGGCCTGATT  | 11460 |
| GGCGGGCACG GGCCCCAGAA CAAGCAGCCC TTCATGGTGG CTTTCTTCAA GCCCACGGAG  | 11520 |
| GTCCACTTCC GCAGCATCCG GTCCACGGGG AGCAAACAGC GCAGCCAGAA CCGCTCCAAG  | 11580 |
| ACGCCAAGA ACCAGGAAGC CCTCGGATGG CCAACGTGGC AGGGTATCTT AGGTGGGAGG   | 11640 |
| GATCACAGAC CCACCACAGG AACCCAGCAG GCCCGGGCGA CCGCAGGAGA CTGACTAAAA  | 11700 |
| TCATTCACTG CTCACCAAGA TGCTCTGAGC TCTCTTCGAT TTTAGCAAAC CAGGAGTCCG  | 11757 |
| AAGATCTAAG GAGAGCTGGG GGTTTGACTC CGAGAGCTCG AGCAGTCCCC AAGACCTGGT  | 11820 |
| CTTGACTCAC GAGTTAGACT CCACTCAGAG GCTGACTGTC TCCAGGGTCT ACACCTCTAA  | 11880 |
| GGCGACACT GGGCTCAAGC AGACTGCCGT TTTCTATATG GGATGAGCCT TCACAGGGCA   | 11940 |
| GCCAGTTGGG ATGGGTTGAG CTTTGGCTGT AGACATCAGA AACCCAAGTC AAATGCGCTT  | 12000 |
| CAACCAGTAG AAAATTCAACC AGCCCGCAGA GCTAAGGTTG GGTGGACATT AGGGTTGGTT | 12060 |
| GATCCAGGAG CTCAACAGTG TCCTCTGAGC CCCAGCTCCT TCTGCCAAC CCCACCATCT   | 12120 |
| TCAGTGCTGC TTCCCTCTCAA GGCCACAGCT GTAGTTGGCC AGGGGGGCTT CATTATTTTT | 12180 |
| TGCTCCTGGG CAGTAGGAGG AAGAGAATGA ATGTCTCTCC ATGGGTCTTT CTTAGGAATG  | 12240 |
| TGGGAACCTT TTCCAGAAAGT CTCTATGTCT TTTAGTTGT GTTGGGTCAC TTGCCCTTCC  | 12300 |
| TGAACCACTT CCTGACTCCT GGACAGGATG TGCACTGATG ACCTTAGCTT TGGGGATCTA  | 12360 |
| ATAGTGACTT TACAAAGCCT CTTTGAGAAG GTGACATTGG AACCAAGGCT TGAGCAGACA  | 12420 |
| CAACAAAGAT TGCAGGGAGG GGCATTGCAAG GTGGAGGAAA CGGCACATGC AAGAGCCCTG | 12480 |
| CGTGGGAGTG AGCTTGGTGT TTGGTCAATC AGTTGTCAGA GCACACCGGG CCCTGTCAGC  | 12540 |
| AGGCACAGCC TGGGCCTGCT CTGAGTATGA CAGAGAGCCC CTGGGAAGTT GTAGGTGGAG  | 12600 |
| GAAAGACAGG TCATGACTAG GAAAAAAAGCA ATCCCTCTGT TGTGGGGTGG AAGGAAGGTT | 12660 |
| GCAGTGTGTG TGAGAGAGAG ACAAGACAGA CAGACAGACA CTTCTCAATG TTTACAAGTG  | 12720 |
| CTCAGGCCCT GACCCGAATG CTTCCAAATT TACGTAGTTC TGGAAAACCC CCTGTATCAT  | 12780 |

|   |       |
|---|-------|
| TTTCACTACT CAAAGAAACC TCGGGAGTGT TTTCTTCTGA AAGGTCACTCA GGTTTTGACT  | 12840 |
| CTCTGCTGTC TCATTTCTTC TTGCTGGTGG TGGTGTGGT TGCTTGTCCC AGGCCCTGTC    | 12900 |
| CCGCATCCTC TTGCCCCCTGC AGAGGGATGA GTGTGTTGGG GCCTCACGAG TTGAGGTTGT  | 12960 |
| TCATAAGCAG ATCTCTTGA GCAGGGCGCC TGCAGTGGCC TTGTTGAGG CTGGAGGGGT     | 13020 |
| TTCGATTCCC TTATGGAATC CAGGCAGATG TAGCATTAA ACAACACACG TGTATAAAAG    | 13080 |
| AAACCACTGT CCCCAGAAGG TTCCAGAAAG TATTATGGGA TAAGACTACA TGAGAGAGGA   | 13140 |
| ATGGGGCATT GGCACCTCCC TTAGTAGGGC CTTTGCTGGG GGTAGAAATG AGTTTTAAGG   | 13200 |
| CAGGTTAGAC CCTCGAACTG GCTTTGAAT CGGGAAATT ACCCCCCAGC CGTTCTGTGC     | 13260 |
| TTCATTGCTG TTCACATCAC TGCCTAAGAT GGAGGAACCT TGATGTGTGT GTGTTCTTT    | 13320 |
| CTCCTCACTG GGCTCTGCTT CTTCACTTCC TTGTCATGC AGAGAACAGC AGCAGGCACC    | 13380 |
| AGAGGCAGGC CTTGTAAGAA GCACGAGCTG TATGTCAGCT TCCGAGACCT GGGCTGGCAG   | 13440 |
| GTAAGGGGCT GGCTGGGTCT GTCTTGGGTG TGGGCCCTCT GGCGTGGGCT CCCACAGGCA   | 13500 |
| GCGGGTGCTG TGCTCAGTCT TGTTCTCAT CTCTGCCAGT TAAGACTCCA GTATCAAGTG    | 13560 |
| GCCTCGCTAG GGAAGGGGAC TTGGGCTAAG GATACAGGGA GGCTCATGA AATCCGAGAG    | 13620 |
| CAGAAATGTG GTTGAGACTT GAACTCGAAC CAGGAACCCA AACACTTTGG ACTCTGAACC   | 13680 |
| CCATTCTCTG CATGCACCTC ATTCCCATCC CTTGGCTGGC TGCTTCTCAA GATGATGCCG   | 13740 |
| GGCCGTGTGT TTGAATGTAG ATACCTGGGG AGCCATCTCC CCCTCTGCC CTCGACTTCA    | 13800 |
| TTTACCCAT TCCCATTCCC ACGGGAGGGA CGGATCTCCC CAGCTTGGTT CAGGCGCTTG    | 13860 |
| TTCCTGAACC AGTCAACTGT TTCAGGGGTG GGGTCATGTT ACTGGCACAT GGCTGCCCC    | 13920 |
| TCTGGAGCCA TTTGCATGGA GTGAGGCAAAGGCAAGGGGA TGAATCTAGG AGAGGAGTGA    | 13980 |
| GGGTTCATGTG ATCCACCTGC CGTGAGCTCT GGATCGTGAT TCTCATTCAAG CAGTCACGAG | 14040 |
| CATCTCGAGC GTTCTGGGCC CTGTTCTAGG TACTGGATTG GAGATGCAGC GATGAACACT   | 14100 |
| GCAATGTGTC TGCCCTGTGG GGCTCAAATA TCCCTGGAGA GGGTATTGTC ATGAGGTCA    | 14160 |
| CAGGGCAACT GGTGGTATTC TACCCCTCAGG GAGCTTGTAG TTCAGTGGGA GAGTCCAGAA  | 14220 |
| TCTTCCCTGG GGATTATGCC CAGACACACT CAGGGCGTAC GTGCACACAG CCAGCTCTGA   | 14280 |
| GCCCTCCTGT GAGCCTGCC TCAGGACTGA TGACCACATC TACCTGCAGC TGGGACAGAA    | 14340 |
| CCCAAACTCC AGGGGCCTCT GCTGGAAGAT TCCATGTGCT TAAGCATCAC TGAGGAGTAT   | 14400 |

|  |       |
|--|-------|
| ATTGATTATT GGGCAACATT TCTGTGCCAC CCAGACCTA GAGGCAAGGA TGGCACATGG   | 14460 |
| ATCCCTTACT GACCAGTGCA CCCGGAGCCA GCATGGGTGA TGCCATTATG AGTTATTAGC  | 14520 |
| CTCTCTGGCA GGTGGGCAAA CCGAGGCATG GAGGTTGTT TAAGGTGAAC TGCCAGTGTG   | 14580 |
| TGACCACCTA GTGGGGTAG AGCTGATGAT TGCCACAC CGGAGGCTCC TTCCGTGCC      | 14640 |
| CGCTTCTGTC CAGAAGACAC AGCCATGGAT GTCCATTAA GGATCAGCCA AGCCCGTGGG   | 14700 |
| GCTTCCTTC ATTTTATTT TATGTTTT TAGAAATGGG GTCTGCTCT GTCACCCAGG       | 14760 |
| CTGGGGTGCA GTGGTGTGAT CATACTCAC CGCAGCTTG AGCCGCTTC CCACTCAGTC     | 14820 |
| TACTAAGCTT GGACTATAGG CCAAGACTAT AGAGTGGTCC TTCTTCCAT TCTTTGGGA    | 14880 |
| CCATGAGAGG CCACCCATGT TTCCGTCCCC TGCTGGGCC TGCTGCTCAG AAGGCATGGT   | 14940 |
| CTGAGGCTTT CACCTTGGTC GTGAGCCTTC GTGGTGGTTT CTTTCAGCAT GGGGTTGGGA  | 15000 |
| TGCTGTGCTC AGGCTTCTGC ATGGTTCCC ACACTCTT CTCCTCCTCA GGACTGGATC     | 15060 |
| ATCGCGCCTG AAGGCTACGC GCGCTACTAC TGTGAGGGGG AGTGTGCCTT CCCTCTGAAC  | 15120 |
| TCCTACATGA ACGCCACCAA CCACGCCATC GTGCAGACGC TGGTGGGTGT CACGCCATCT  | 15180 |
| TGGGGTGTGG TCACCTGGGC CGGGCAGGCT GCGGGGCCAC CAGATCCTGC TGCCCTCCAAG | 15240 |
| CTGGGGCCTG AGTAGATGTC AGCCATTGC CATGTCATGA CTTTGCGGGG CCCCTGCGC    | 15300 |
| CGTTAAAAAA AAATCAAAAA TTGTACTTTA TGACTGGTTT GGTATAAAGA GGAGTATAAT  | 15360 |
| CTTCGACCCCT GGAGTTCATT TATTCTCCT AATTTTAAA GTAACTAAAA GTTGTATGGG   | 15420 |
| CTCCTTGAG GATGCTTGT A GTATTGTGGG TGCTGGTTAC GGTGCCTAAG AGCACTGGGC  | 15480 |
| CCCTGCTTCA TTTTCCAGTA GAGGAAACAG GTAAACAGAT GAGAAATTTC AGTGAGGGC   | 15540 |
| ACAGTGATCA GAAGCGGGCC ACCAGGATAA TGGGATGGAG AGATGAGTGG GGACCCATGG  | 15600 |
| GCCATTTCAA GTAAATTTC AGTCGGGTCA CCAGGAAGAT TCCATGTGAT AATGAGATTA   | 15660 |
| ACGTGCCAG TCACGGCGAC ACTCAGTAGG TGTATTCCCT GCTCTGCCAA CAGCAACCCT   | 15720 |
| AGTTGATAAG AGCTGTTAGG GATTTGTCC TTTTGCTTAG AATCCAAGGT TCAAGGACCT   | 15780 |
| TGGTTATGTA GCTCCCTGTC ATGAACATCA TCTGAGCCTT TCCTGCCTAC TGATCATCCA  | 15840 |
| CCCTGCCTTG AATGCTTCTA GTGACAGAGA GCTCACTACC AGGACTACTC CCTCCTTCA   | 15900 |
| TTTAGTAATC TGCCTCCCTC TTTCTTGTC CCTGCTCTGT GTGTTAAGTC CTGGAGAAAA   | 15960 |
| ATCTCATCTA TCCCTTCAT TTGATTCTGC TCTTGAGGG CAGGGTTTT TGTTCTTG       | 16020 |

|            |            |            |            |            |            |       |
|------------|------------|------------|------------|------------|------------|-------|
| TTTGTTCCTT | TAAGTGTGG  | TTTCCAAAG  | CCCTGCTCC  | CCTCCTCAAT | TGAAACTTCA | 16080 |
| AAGCCCTCAT | TGGGATTGAA | GGTCCTTAGG | CTGGAAACAG | AAGAGTCCTC | CCCAACCTGT | 16140 |
| TCCCTGGCCT | GGATGTGCTG | TGCTGTGCCA | GTATCCCCTG | GAAGGTGCCA | GGCATGTCTC | 16200 |
| CCCGGCTGCC | AGGGGACACA | TCTCTATCCT | TCTCCAACCC | CTGCCTTCAT | GGCCCATGGA | 16260 |
| ACAGGAGTGC | CATGCCCTG  | TGTGCACCTA | CTTCCATCAG | TATTCACCA  | GAGATCTGCA | 16320 |
| GGATCAAAGT | GAATTCTCCA | GGGATTGTGA | AATGATGCGA | TTGTGGTCAT | GTTTAAAAGG | 16380 |
| GGGCAACTGT | CTTCTAGAGA | GTCCTGATGA | AATGCTTCCA | GAGGAAATGA | GCTGATGGCT | 16440 |
| GGAATTGCT  | TTAAAATCAT | TCAAGGTGGA | GCAGGTGGGG | AAGGGTATGG | ATGTGTAAGA | 16500 |
| GTTCGAAATT | GTCCATCATA | AAATGTGTA  | AAAGCATGCT | GGCCTATGTC | AGCAGTCACA | 16560 |
| GCCTGGAGGT | GGTAACAGAG | TGCCAGTCAC | TGATGCTCAA | GCCTGGCACC | TACAGTTGCT | 16620 |
| GGAAACCCAG | AAGTTTCACG | TTGAAAACAA | CAGGACAGTG | GAATCTCTGG | CCCTGTCTTG | 16680 |
| AACACGTGGC | AGATCTGCTA | ACACTGATCT | TGGTTGGCTG | CCGTCAGCTT | AGGTTGAGTG | 16740 |
| GCGGTCTTCC | CTTAGTTTGC | TTAGTCCCCG | CTATTCCCTA | TTGTCTTACC | TCGGTCTATT | 16800 |
| TTGCTTATCA | GTGGACCTCA | CGAGGCACTC | ATAGGCATTT | GAGTCTATGT | GTCCCTGTCC | 16860 |
| CACATCCTCT | GTAAGGTGCA | GAGAAGTCCA | TGAGCAAGAT | GGAGCACTTC | TAGTGGGTCC | 16920 |
| AAGTCAGGGA | CACTATTCA  | CAATCTACAG | TGCACAGGGC | AGTTCCCCAA | CAGAGAATT  | 16980 |
| CCTGGTCCTG | AATGTCGGAT | CTGGCCCCTT | CCTTCCCCAC | TGTATAATGT | GAAAACCTCT | 17040 |
| ATGCTTTGTT | CCCCTTGTCT | GCAAAACAGG | GATAATCCCA | GAACTGAGTT | GTCCATGTAA | 17070 |
| AGTGCTTACA | ACAGGGAGTG | CTTGGCTTGG | GGAGTGTAC  | CTGCAGTCAT | TCATTATGCC | 17160 |
| CAGACAGGAT | GTTCCTTTAT | AGAAACGTGG | AGGCCAGTTA | GAACGACTCA | CCGCTTCTCA | 17220 |
| CCACTGCCCA | TGTTTGGTG  | TGTGTTTCAG | GTCCACTTCA | TCAACCCGGA | AACGGTGCC  | 17280 |
| AAGCCCTGCT | GTGGCCCCAC | GCAGCTCAAT | GCCATCTCCG | TCCTCTACTT | CGATGACAGC | 17340 |
| TCCAACGTCA | TCCTGAAGAA | ATACAGAAC  | ATGGTGGTCC | GGGCCTGTGG | CTGCCACTAG | 17400 |
| CTCCTCCGAG | AATTC      |            |            |            |            | 17415 |

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2298 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: misc\_feature

(B) LOCATION: 1..2298

(D) OTHER INFORMATION: /note= "MOP1 UPSTREAM SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

|  |      |
|--|------|
| TGCATAGGTC ACACATCCCT CCTCTACCCA AGGCTAGCCA GGTGCCCTAT CTCTCCCTTC  | 60   |
| TTCGGTGCCT CCTCCGACTG GGCTCTGACG TTCTCAGAGA GAAAGAAAGG GAAAGACTGC  | 120  |
| TTGCTACCCCT TCTGTTCCGG ACCTTACTGA AGGGCCTTAG TGTTTCCAGG GGCCCAAGAA | 180  |
| CCAGGTAGCC GTGGAGGTTG CCATGCTTGC CTGCCCACTC ACCCAACGTT CTCCCTGCCTG | 240  |
| GCCTGTGTGT GGCACCCATG CAGGCCACAG AAGGCCACAC ACAGCCTTCA GGATGAGGCA  | 300  |
| GGGCCCCTTT GTTTATTCAA TATCAAGAAC TGTAACGTGG TCACCGGAGG TCATGTCTCC  | 360  |
| AGCTCGCAGC CTGCTTGGCC TCAGATCACC CACCACAGCA GGTCCAGGGA GGGGCCTCTC  | 420  |
| AGGTCTGCAC TGGGCCAGGG ACTCAGTACT GGTGGGCATC CAAGGCCTGG GCTAAGACCT  | 480  |
| GCAAGTTCT TTTAGCCCCCT CAGACAGTCA CATCACCTAA AATTCCCTACC AAGGAGCCCT | 540  |
| GAGAGACCTA GGTAGTTATC TCTGTTCCAG GAAGCCTGAA AGACCAGGCT TCCCATCTCA  | 600  |
| CCCTAGGACT TCAAGAGGGGA CCCCCCTACTC AAGGCCCTTC CCCAGCCCT ACTTGCCATT | 660  |
| TTACCACCCC TGAAACGCTT GCTTGTGCC CACCTTCAGC AAAGCAGGAA GCCTGGCTCA   | 720  |
| CCATCCCCAC TCACTCACTG CCATTCTGGG TGAAGGCTGC TTTGCTCCCA TTTTCAGAT   | 780  |
| TAGGAAACGG AGGCTCCAAA GAGCAGCAAT CCACTGAGAG ACCCAGTATC TGTCTGGAC   | 840  |
| GTTCCTCCT GGGAGGAGAG GGAGGCTAGT CCTTGAGAC AGGAAAATCG AGTCGGGAGC    | 900  |
| TCTTCTGAAC TTGGGTACCA ACTGCCTACT CCTCAGGCC CTGACCTGGG GCTAGGGGTA   | 960  |
| GGGGTTATTA GACAGTGAGG TACCAAAGGA CTCATGTCAG GACCCGGCCC CCCCCAAGAGA | 1020 |
| GGAGGGGGTG GGAACATTCT CTAGTCCCAG ATTCACCTTA TGTACTCTGT AGAGCTGCAG  | 1080 |
| CATCTGGGGT TTGAAGGCTT TGGTTAAAA GATACTTGGG AAGGAAAAGC CGAGAAGTAC   | 1140 |
| CTGGGCCCGG ATCCCTTGGG TGCTGGACTT GAGGGGAGGT GTGTGTGTGT GTGTGTGTGA  | 1200 |
| GTGTGTGTGT ATGTATGTGT GTGTTGGGG AGTGAAGTGT AGAAAGAACT TTATCTCCAC   | 1260 |

|   |      |
|---|------|
| ATTATCTCTG CCCGTCTGG AAGGTTCCA GAGGAAGTGG CACCCGAGGG GGGAGGGGCA   | 1320 |
| GGGAGAACGT TCCCCGAGGA ACAAAAGCCA GGATAGCAGA GGGGCAAGCG GTGGGGTAC  | 1380 |
| CGAGGGGTT TTGCATGACT GGAGCAAATG GAGTGTGAGG GGGGGCGGTT CGAAAGATGA  | 1440 |
| GCCAGGTCCA AGAGTGGCCA CCTCCGAGGA GCCTCTCGG ATTCCTGCCG TCCCTCCTGG  | 1500 |
| ATGCTTCCT AGCACAGCCC TTAGTTGCTA CACTTGGCC ACTTCCAAGT GCGAGTCCCG   | 1560 |
| AGAGAGCTGG GCAGATTGGG ATTCTTCTCT CTGGGTCCCT GCGGGCTCTG TCCCAGTGCC | 1620 |
| GGACACCCGG TGGCACTCG GTAAATATTT GTAGAGCGCC CTGGGAGGAA TGAATGAAGC  | 1680 |
| CATTGGGCCA GGCTTGGGA GGGCGGGAC AGGCGCAGGT GGGAGGCAGC GGGAGCGGGA   | 1740 |
| GGGGCGGGGA AGTCAGTCCT CCCGCTCCTC CCCCCTCAGC CGGCCCAACT            | 1800 |
| CCGGGGCTCC CGAGGGCGCG GGCGGCGAT CCGGGCGCGC AGGGCCCTTG TATTGGCAC   | 1860 |
| GCGGGAGATC GGAAAGGGGT TTGTTGCTGG TGCCCGCGG CCTGAGCGCG ATCAGAGCGG  | 1920 |
| GAGGAGGGAG CTAGGGTTCG CTCAGGCCCG AGCTGCCTCT CGGGCACTCG CTCTCCGGAC | 1980 |
| TGTAGGTCTG CAAGCTGCTG CTCCCTCCAC CCCGGCCCGC CTCCTCGCTC TCTTGCTCGC | 2040 |
| TCTCTGGAGT TGCTGTGCTA GCCTTGCCGT GCGTCCTGGC GACTGCGGGC CGAGGGGCC  | 2100 |
| CGGGCCAGAA CTGAGTAAAG GACAGGGCG TCCCGGGCAA AGCGCAGCCG GCGGGGGAGT  | 2160 |
| GGCCATGTGT GGCGAGGCCG CCTTGAAGCT CGCCTGCAGC AAGTGACCTC GGGTCGTGGA | 2220 |
| CCGCTGCCCT GCCCCCTCCG CTGCCACCTG GGGCGCGCG GGGCCGGTGC CCCGGATCGC  | 2280 |
| CGCTAGAGCC GGCGCGATG  | 2299 |

## (2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2997 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 1..2997
- (D) OTHER INFORMATION: /note= "MOP1 TERMINAL SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

|             |             |                     |            |            |      |
|-------------|-------------|---------------------|------------|------------|------|
| TAGCTCTTCC  | TGAGACCCCTG | ACCTTTGC GGCCACACCT | TTCCAAATCT | TCGATGTCTC | 60   |
| ACCATCTAAG  | TCTCTCACTG  | CCCACCTTGG          | CGAGGAGCCA | ACAGACCAAC | 120  |
| CCTTCCCCTC  | ACCTCCCCAA  | CCGGAAGCAT          | GTAAGGGTTC | CAGAACCTG  | 180  |
| CAGCTGATGA  | GCGCCCTTTC  | CTTCTGGCAC          | GTGACGGACA | AGATCCTACC | 240  |
| GCAAACGCC   | AAGAGCAGGA  | AAAATGTCTG          | CCAGGAAAGT | GTCCATTGGC | 300  |
| CTGGCGCTCT  | GAGTCTTGA   | GGAGTAATCG          | CAAGCCTCGT | TCAGCTGCAG | 360  |
| GGCTTAGCCA  | GGGTGGGCGC  | TGGCGTCTGT          | GTTGAAGGAA | AACCAAGCAG | 420  |
| AATGATATGT  | CACAATAAAA  | CCCATGAATG          | AAAATGGTTA | GGATACAGAT | 480  |
| AAACAATTAA  | TCCCCGTTTC  | TTGGTTTATT          | CTGACTTTGT | AAACAGAAAA | 540  |
| TGGAGGATGG  | AGAGGCCCT   | CCTTCCGTC           | TCGTCTCGTT | GTGTGTGTTT | 600  |
| CCCAAATCCA  | GCCTGTAGGG  | AGGAGGAGGA          | GGATGTCTGC | TCAGAAGAGG | 660  |
| ATGTGGCCTC  | AAAGGGTGT   | GGGATGAAGA          | TGGAGGGAGG | TATGCATGCA | 720  |
| CACACACACA  | CACACACACA  | CATGCATGAT          | ACACACACAC | ACACACACAC | 780  |
| TGCACACACA  | CACACACACA  | CACACACACA          | CACACACGCA | CGCACGCACG | 840  |
| CATGCATGCA  | CACACACACG  | CACACACACA          | TCTGAAGCGC | ATGTAGACTT | 900  |
| CTGCCAGTCC  | CTCAGCCCCA  | ATTCCCTGCC          | CATGGTAGGA | AATCCATGAG | 960  |
| CTAACAAAGCA | CAGCGGACCC  | TACCTGAGGA          | ACCACAGGGG | ATGCAGGCTC | 1020 |
| TGTCCTCCAA  | ACAAGGCCCC  | TCTGGCACCT          | CTGTGGCCGA | GCTCCGGAGC | 1080 |
| CCTTCACAGC  | TGCCTCTCTT  | CACTCTCAAC          | CCTAACAGAA | GGTTCTGCGA | 1140 |
| TCTGGATCTG  | AGGGAGATGG  | CAGAACAGGG          | TTGTACTGGC | TTAGAAGGTT | 1200 |
| TCCTGCTTCA  | GAGGTGGGAT  | GTTGGTTATG          | GCTCAAACAA | GGCCTCTCTG | 1260 |
| CAGAGCCCCA  | GCTGCCCAA   | TGGTTCTAG           | CTTCAAATGC | AGAGGGTTAA | 1320 |
| AGTGTTCCT   | GCATCCACAC  | AAAGAATGAG          | GTTAGCCAGG | CAGGACCTAT | 1380 |
| CATCTGGTCA  | GGTGGGAAAC  | CAATTCTCA           | TGTCTGTGTC | CCTGGAAACA | 1440 |
| TTCTGTTCTG  | TTTAGTTTT   | TCTTCTTCAG          | TAGCTTGGGC | TGCAGCTTCT | 1500 |
| TTCGATGTGG  | GGGAAGGCCA  | TTCTTTTG            | TAATTTGTTC | TGTGTGTTG  | 1560 |

|             |             |            |             |             |            |      |
|-------------|-------------|------------|-------------|-------------|------------|------|
| GCTTTTGTCG  | TGACTCCCCT  | GTGGTGCACA | TTTTACTTTA  | GAGCCCTAGT  | CTGCCTGCAG | 1620 |
| TCGGTGTCTC  | TTATACGTTT  | AAATGTGTAA | ATAGTTGTGA  | CAAGACAAAG  | AAATTATTTA | 1680 |
| TTTCCATCTG  | AAGCTTTTC   | CAAAGGCTCC | TCACAGAGAA  | CAATGAGGCC  | GACTTCCTTC | 1740 |
| AGTCTGTTG   | TTTCTTATT   | TAAGACTATT | TATTAACAGT  | TGGACCGATG  | TACCCATAGC | 1800 |
| TGTCGAATAA  | AGTGGTCCTT  | AGTAAAATT  | CTGTATAAAT  | AGAGTAAGAA  | GGGGTTTGAC | 1860 |
| TTTGCAATAA  | AAGGAGACAT  | TTGGTTCTGG | TTGTCCGACC  | CATGTGTGTA  | TTTGTGTCTT | 1920 |
| TCCCCCTGAA  | CTCCTGGACA  | CTGGAGTCTC | ATCGGCTGAG  | AACCCTCGAC  | CTTGATCTCG | 1980 |
| ACTGTTAACG  | GGATGTTTAT  | CATCCAGGCC | CAGGGGAAGT  | CGGGCGCTCC  | TCAATATTTG | 2040 |
| GTGCAGCTGT  | GTGGGGCTCC  | CTGGGCGGGA | GAGACGGAAC  | CAAACAAACAA | ATGTGAGTTT | 2100 |
| GGTAAGGCTG  | GATGGCAAAG  | AGTGCCTTG  | ATTGAACATAC | AGCCCAGCTG  | TCAGCAGCTG | 2160 |
| CTTCAAAGAG  | GCAGGGGGTA  | AATTAGCTGT | GTAACTGCT   | AACATAGTCG  | AAAGATTTAG | 2220 |
| TCATCCCAAT  | AAAATAGAGG  | CACAAGAGAG | AAGAGGGGGG  | GGTGTATACC  | CCAAACTTGA | 2280 |
| AAGCCATGCT  | GGCCTCACAG  | CTGGCGTCAT | TCAGTGCCCCG | TCACACCCGG  | GCAGTTGGGG | 2340 |
| GCTGCCCTCG  | CAGGCCAAGC  | TGTGGAGGTG | GGCAGCCCCAC | CGCAGGCTGG  | AGAACGGAGT | 2400 |
| GCCCCCCCACC | TCCCCGGCAA  | GCTCAGGGCA | GTGCTCATCT  | GGCTACATCG  | GTCTTGAAAG | 2460 |
| TGCGCACGAA  | GGTCACCTGA  | CGGATGTTTC | TAGAATCCCA  | GGCGATGCTT  | GGGACAGGCT | 2520 |
| GCTCTCTCTT  | CCCCGTGTTGA | CTCAGACCCA | GCAACCCAGC  | CGTCCTAACCA | CATTCCAGCC | 2580 |
| CCTGCGATTT  | CTAAACCTTT  | CCTGTCACTG | TCCCGACAAC  | TCAGCTTTG   | TTCTGTTTTC | 2640 |
| CAGGCTGAAG  | CCCAGAGCCA  | CAAGCCGGAG | GGTCCAGATG  | TGGCCTCTCA  | GATGTGTGCC | 2700 |
| TTAGCCTCTC  | AACCCCACCC  | CCACCCCCAA | CCCCAGTGAT  | GTTTACACAT  | CTTAAAAAAC | 2760 |
| ACTAATCTGT  | TGCCAATATG  | TTTTTGCAAA | TAAGGACTTT  | GGGCTTCTCT  | TGAGCGGGCC | 2820 |
| ACCTGGTTCC  | TCCCTGTGTG  | CTGCTCCTAA | CTGAACAGAG  | GTGCCAGGGC  | CGTTGTCACA | 2880 |
| CATACACACA  | CCCCCGCCAT  | GGCCTCATCC | ACAAACGGTC  | GAGGTCAGCT  | GACATCTCA  | 2940 |
| AAATGGCTGA  | CGGATGTCTA  | CTTGTGCCCA | CGACCCAAAA  | GGAATAGGAA  | AATGGAA    | 2997 |

## (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 9 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 1..9
- (D) OTHER INFORMATION: /note= "WT1/EGR CONSENSUS SEQUENCE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GNGNGGGNG

9

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 1..21
- (D) OTHER INFORMATION: /note= "WT1/EGR HUMAN TCC BINDING SITE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TCCTCCTCCT CCTCCTCCTC C

21

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

- (A) NAME/KEY: misc\_feature
- (B) LOCATION: 1..15
- (D) OTHER INFORMATION: /note= "WT1/EGR MOUSE TCC BINDING SITE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:  
(A) NAME/KEY: misc\_feature  
(B) LOCATION: 1..9  
(D) OTHER INFORMATION: /note= "HUMAN FTZ BINDING SITE"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TCAAGGTCA

9

What is claimed is:

1. A vector comprising a DNA sequence defining a reporter gene in operative association with at least one OP1-specific non-coding sequence lying contiguous to the OP1 gene under naturally-occurring conditions and competent to affect expression of said reporter gene on said vector.
2. The vector of claim 1 wherein said non-coding sequence is capable of being acted on by a nucleic acid binding molecule, thereby to affect expression of said reporter gene.
3. The vector of claim 1 wherein said non-coding sequence is selected from the group of DNA sequences defined by bases 3170 to 3317 (Seq. ID No. 1); 3020 to 3317 (Seq. ID No. 1); 2790 to 3317 (Seq. ID No. 1); 2548 to 3317 (Seq. ID No. 1); 2150 to 2296 (Seq. ID No. 2); 2000 to 2296 (Seq. ID No. 2); 1788 to 2296 (Seq. ID No. 2); 1549 to 2296 (Seq. ID No. 2), including allelic, species and other sequence variants thereof.
4. The vector of claim 1 wherein said non-coding sequence is selected from the group of DNA sequences defined by bases 2300 to 3317 (Seq. ID No. 1); 1300 to 3317 (Seq. ID No. 1); 1 to 3317 (Seq. ID No. 1); 2548 to 2790 (Seq. ID No. 1); 1549 to 2790 (Seq. ID No. 1), 1 to 2790 (Seq. ID No. 1); 800 to 2296 (Seq. ID No. 2); 1 to 2296 (Seq. ID No. 2); 1549 to 1788 (Seq. ID No. 2); 800 to 1788 (Seq. ID No. 2); 1 to 1788 (Seq. ID No. 2), including allelic, species and other sequence variants thereof.
5. The vector of claim 1 wherein said non-coding sequence is defined by part or all of Seq. ID No. 3 including allelic, species and other sequence variant thereof.
6. The vector of claim 1 wherein said non-coding sequence comprises part or all of an OP1 intron sequence.

7. The vector of claim 6 wherein said sequence defining part or all of an OP1 intron is selected from the group of sequences consisting of bases 3736 to 10700; bases 10897 to 11063; bases 11217 to 11424; bases 11623 to 13358; bases 13440 to 10548; bases 15166 to 17250; all of Seq. ID No. 1, including allelic, species and other sequence variants thereof.
8. The vector of claim 1 wherein said vector comprises at least a second said non-coding sequence.
9. The vector of claim 8 wherein said second non-coding sequence is independently selected from the group of sequences defined in claims 2, 3, 4, 5 or 7.
10. The vector of claim 1 wherein said non-coding sequence defines at least one Wt-1/Egr consensus binding element.
11. The vector of claim 1 wherein said non-coding sequence defines between one and six Wt-1/Egr binding elements.
12. The vector of claim 1 wherein said non-coding sequence defines at least part of an FTZ binding element.
13. The vector of claim 1 wherein said non-coding sequence defines a steroid binding element.
14. A cell transfected with a vector of any of claims 1, 3, 4, 5, 6, 10, 11 or 12.
15. The transfected cell of claim 14 wherein at least part of the DNA of said vector is operatively integrated into the cellular genome.
16. The transfected cell of claim 15 wherein said cell's genome has an OP-1 gene locus and at least part of said transfected DNA is operatively integrated into said genome at said OP-1 locus.
17. The transfected cell of claim 14 wherein said cell expresses OP1 under naturally-occurring conditions.
18. The transfected cell of claim 17 wherein said cell is an epithelial cell.

19. The transfected cell of claim 17 wherein said cell is of kidney, renal, urogenital, liver, bone, cardiac, lung, or nerve cell origin.
20. A cell comprising a transfected vector, said vector defining a reporter gene in operative association with at least two DNA sequences,  
the first said sequence comprising part or all of a sequence selected from the group consisting of bases 2548 to 2790 (Seq. ID No. 1); bases 2548 to 3317 (Seq. ID No. 1); bases 1549 to 1788 (Seq. ID No. 2); bases 1549 to 2296 (Seq. ID No. 2), including allelic, species and other sequence variants thereof, and  
the second said sequence defining a sequence capable of being acted on by a DNA binding molecule and competent to affect expression of said reporter gene.
21. The cell of claim 20 wherein said second DNA sequence comprises at least one Wt-1/Egr-1 consensus element (Seq. ID No. 4).
22. The cell of claim 21 wherein said second DNA sequence comprises between one and six Wt-1/Egr-1 consensus elements (Seq. ID No. 4).
23. The cell of claim 21 wherein said second DNA sequence comprises at least six Wt-1/Egr-1 consensus elements (Seq. ID No. 4).
24. The cell of claim 20 wherein said second DNA sequence is selected from the group of sequences consisting of a TCC element, an FTZ binding element and a steroid binding element.
25. The cell of claim 20 further comprising a third DNA sequence in operative association with said reporter gene and competent to affect expression of said gene, said third DNA sequence being independently selected from the group of

sequences consisting of a TCC element, an FTZ binding element and a steroid binding element.

26. The cell of claim 20 further comprising a third DNA sequence in operative association with said reporter gene and competent to affect expression of said gene, said third DNA sequence being independently selected from the group of sequences consisting of bases 3736 to 10700 (Seq. ID No. 1); bases 10897 to 11063 (Seq. ID No. 1); bases 11217 to 11424 (Seq. ID No. 1); bases 11623 to 13358 (Seq. ID No. 1); bases 13440 to 10548 (Seq. ID No. 1); bases 15166 to 17250 (Seq. ID No. 1), including allelic, species and other sequence variants thereof.
27. A method for screening a candidate compound for the ability to modulate expression of OP-1, said method comprising the steps of:
  - (a) incubating a said candidate compound with a cell transfected with a vector comprising a DNA sequence defining a reporter gene in operative association with at least one OP1-specific non-coding sequence lying contiguous to the OP1 gene under naturally-occurring conditions and competent to affect expression of said reporter gene on said vector;
  - (b) measuring the level of reporter gene expressed in said cell; and
  - (c) comparing said level with that of said reporter gene expressed in said cell in the absence of said candidate compound, wherein an increase in reporter gene expression level is indicative of said candidate's ability to increase OP-1 expression in vivo, and a decrease in reporter gene expression level is indicative of the candidate's ability to inhibit OP-1 expression in vivo.
28. The method of claim 27 wherein said non-coding sequence is capable of being acted on by a nucleic acid binding

molecule, thereby to affect expression of said reporter gene.

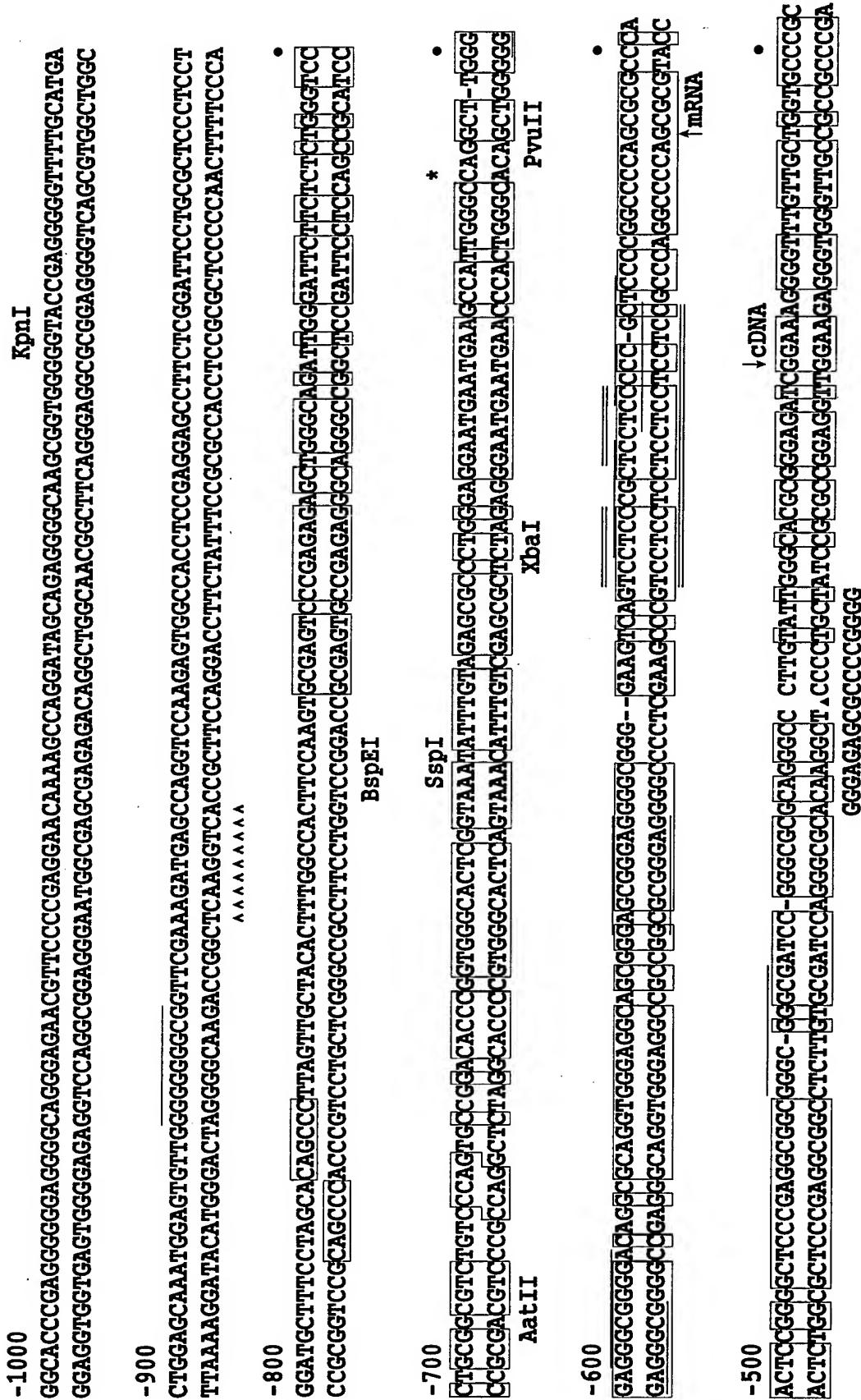
29. The method of claim 27 wherein said non-coding sequence is selected from the group of DNA sequences defined by bases 3170 to 3317 (Seq. ID No. 1); 3020 to 3317 (Seq. ID No. 1); 2790 to 3317 (Seq. ID No. 1); 2548 to 3317 (Seq. ID No. 1); 2150 to 2296 (Seq. ID No. 2); 2000 to 2296 (Seq. ID No. 2); 1788 to 2296 (Seq. ID No. 2); 1549 to 2296 (Seq. ID No. 2), including allelic, species and other sequence variants thereof.
30. The method of claim 27 wherein said non-coding sequence is selected from the group of DNA sequences defined by bases 2300 to 3317 (Seq. ID No. 1); 1300 to 3317 (Seq. ID No. 1); 1 to 3317 (Seq. ID No. 1); 2548 to 2790 (Seq. ID No. 1); 1549 to 2790 (Seq. ID No. 1), 1 to 2790 (Seq. ID No. 1); 800 to 2296 (Seq. ID No. 2); 1 to 2296 (Seq. ID No. 2); 1549 to 1788 (Seq. ID No. 2); 800 to 1788 (Seq. ID No. 2); 1 to 1788 (Seq. ID No. 2), including allelic, species and other sequence variants thereof.
31. The method of claim 27 wherein said non-coding sequence is selected from the group of sequences defined by claims 5, 6 or 7.
32. A method for screening a candidate compound for the ability to modulate expression of OP-1, said method comprising the steps of:
  - (a) incubating a said candidate compound with a cell according to claim 20, 21, 24 or 25;
  - (b) measuring the level of reporter gene expressed in said cell; and
  - (c) comparing said level with that of said reporter gene expressed in said cell in the absence of said candidate compound, wherein an increase in reporter gene expression level is indicative of said

candidate's ability to increase OP-1 expression in vivo, and a decrease in reporter gene expression level is indicative of the candidate's ability to inhibit OP-1 expression in vivo.

33. A compound that is identified by the method of claim 27 or 32.
34. A substantially pure nucleic acid comprising a DNA sequence defined by bases 1 to 1871 of Seq. ID No. 2, including allelic, species and other sequence variants thereof.
35. A substantially pure nucleic acid comprising a DNA sequence defined by bases 1 to 2997 of Seq. ID No. 3, including allelic, species and other sequence variants thereof.
36. The vector of claim 1, 3, 4, 5 or 7 further comprising part or all of a nucleotide sequence encoding an OP1 pro protein in operative association with said reporter gene.
37. The method of claim 27 wherein said vector further comprises part or all of a nucleotide sequence encoding an OP1 pro protein in operative association with said reporter gene.
38. A method for producing a candidate compound having the ability to modulate OP-1 expression in a cell, the method comprising the steps of:
  - (a) obtaining, by the method of claim 27, a candidate compound, and
  - (b) producing either said candidate compound, or a derivative thereof having substantially the same OP-1 expression modulating ability as said candidate.
39. The method of claim 38 wherein said candidate compound, or derivative thereof, produced in step (b) is by recombinant DNA techniques, or by nonbiological peptide synthesis.
40. A kit for identifying a candidate molecule capable of modulating OP-1 expression in a cell, the kit comprising:

- (a) a receptacle adapted to receive a sample, said sample containing a vector encoding a DNA sequence comprising a reporter gene in operative association with at least one OP-1-specific non-coding sequence lying contiguous to the OP-1 gene under naturally-occurring conditions and competent to affect expression of said reporter gene, wherein said vector is carried in a cell,
- (b) means for detecting expression of said reporter gene following exposure of a said candidate compound to said sample.

41. The kit of claim 40 wherein said reporter gene comprises an OP-1 DNA sequence.



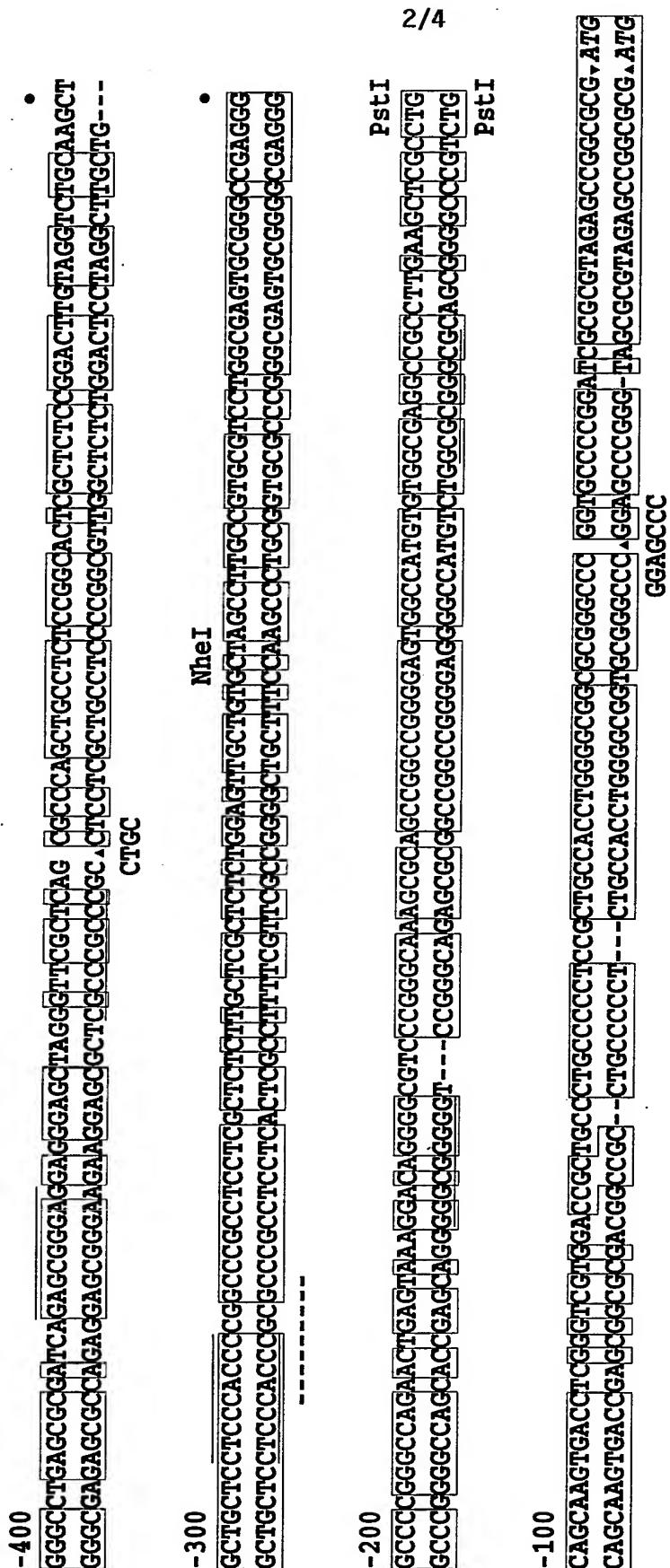
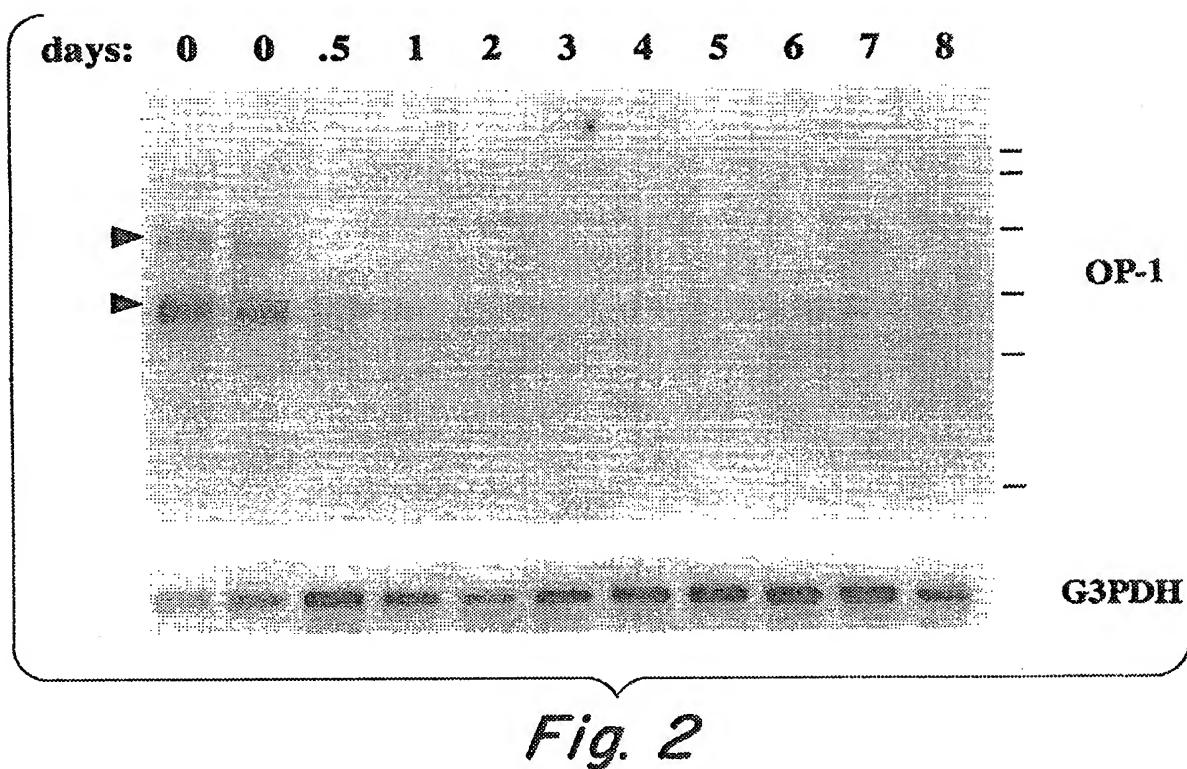
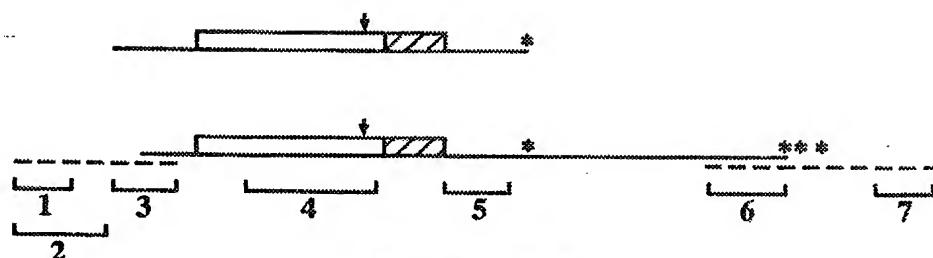


Fig. 1B

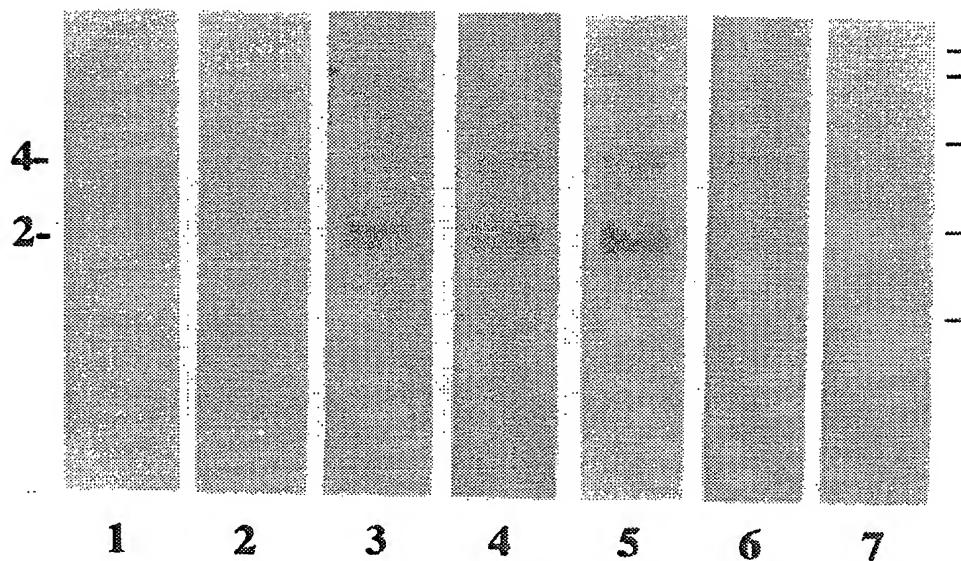
3/4



4/4



*Fig. 3A*



*Fig. 3B*